

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**


```

563 ntrpdygluGlutThrSerSerAsnValProArgProGlyGlnPheArgI 580
458 .....ACAACAACGTGTTCATTCATTCATTCATTCATTCATTCATTCAC 499
580 leproniSerLeuThrMetValProAspGlnGlyGlnLeucysValAla 596
500 AAGTTGATCGTCATTCGAGTACAAATGCTTTTACATGGAAGCTGATAA 549
597 AspArgGluAsnGly...ArgIleGlnCysPheHisAlaGluThrGlyAs 612
550 AACACTGTGTCACAGATTGAGTATCTGAATC...ACAACGTCTTTC 596
612 nPheValLys...GlnIleLysHisGlnGluPheGlyArgGluValPheA 628
597 AACCTCAATGTGCCGATGCCAGTATGCCGTTATGAATTTTGATGCT 646
628 lValSerTyrAlaPro.....Gly 634
647 GGACCAACCGGTCAACCAAGTTCATTCATTCATTCATTCATTCATTCAT 696
635 Gly.....ValLeuTyrAlaValAsnGlyLysProGlyTyr 646
697 TCATATAATGAGCATGCGATTCGAAACGGTATCTTCTGCGGGTGTG 746
646 r.....GlyTyrSerAlaProV 652
747 TCCATTCCTGCTTGTGTCATGTCATGTAACGGTATGTCATGTAATTCCTA 796
652 alGlnGlyPheMetLeuAsnPheSerAsnGlyAsp..... 663
797 AATGCTGATGATGTCCTCTGATTAATTTTG.....CTAAATATTT 840
664 .....IleLeuAspThrPheIleProAlaArgLysAsnPh 675
841 GGAATATCCACAGATTATAGCGCGCCACAGAGTCGATTCGAATCAAT 890
675 eAspMetProHisAspIleAlaAlaAspAspGlyThrValTyrValG 692
891 ATGCGGATGATCAGACGTTTCTATCAATGCCAGTCACT..... 931
692 lAspAlaHisAlaAsnAlaValTTrpLysPheSerProSerLysAlaGlu 708
932 ...ATTACCATTAAGACCAAT.....AGCGATTCGTCGACC 969
709 HisArgSerValLysLysAlaGlyIleGluValGlnGluIleThrGlu 725
970 ACAATGTTCAAGACCAAGATTCGAGCTGTTAAACAGGCGCGCG 1019
725 rGluIlePheGluThrHisIleArgSerArgProLysThrAsnGluSerV 742
1020 CAGCAAAACCTGTCAGCTGCGCAACTTCGTTTACTCAAGAAAGATCT 1069
742 alGlnLysGlnThrGlnGlnLysGlnGlnLys.....GlnLysAsnSer 756
1070 GCA.....GACCGGAGATATCATTT...GATGTACGAC 1101
757 lAlaGlyValSerThrGlnGlnLysGlnAsnValAlaGlnGluIleAsnAl 773
1102 TCAATACACACCTTGAA.....ATTAGCGATGATTAATCAAGCTT 1142
773 aGlyValProThrGlnGlnLysGlnAsnValAlaGlnGluSerSerAlaG 790
1143 TCGCAGTATTCATTCACCGCTGACCGCTCTGCAACATATGACAACT 1192
790 lValSerThrGlnGlnLysGlnSerValAlaGlnGluSerSerAlaGly 806
1193 GTAATPACTTGTCTGACTACAAATGGAATCTGCATCTCACCATTTGGCTT 1242
807 ValSerThrGlnGlnLysGlnSerValAlaGlnGluSerSerAlaGly 823
1243 CTCATGTTTATGGCTTAAGCATTCGATTCATT...GCTGCCGTACATTA 1289

```

```

823 lSerPheValLeuIleIleThrLeuLeuIleIleProIleAlaValLeuI 840
1290 TTACCATTCGTTAAATTCCT 1312
840 leAlaIleAlaIlePheIleArg 847

```

seq_name: A_Geneseq_36:P81184

seq_documentation_block:

```

ID P81184 standard; Protein; 1594 AA.
AC P81184;
DE 29-OCT-1992 (first entry)
DE Sequence encoded by the 2nd reading frame of the peplomeric gene
DE of FIVP strain 79-1146.
KW Vaccine; peplomeric protein; E2 gene.
OS Feline infectious peritonitis virus.
PN EP-264979-A.
PD 27-APR-1988.
PF 01-SEP-1987: 201657.
PR 05-SEP-1986: NL-002244.
PA (DUIIN ) DUPEHR INT RES BV.
PI De Groot RJ, Spaan WJM, Van Der Zeijst BAM,
DR WPI; 88-114147/17.
DR N-PSDB: N81533.
PT Gene for feline infectious peritonitis virus - and gene prod.
PS Useful as antigenic protein for vaccine
PS Disclosure; Fig 1; 13pp. English.
CC cDNA was prep'd. from FIVP strain 79-1146. N81533 gives the sequence
CC of the peplomeric gene in three reading frames. The top reading
CC frame is an open reading frame of 4356 nucleotides and has a coding
CC capacity for a precursor polypeptide having a mol. wt. of 160,470
CC (1452 AAs). The beginning and the end of the E2 gene are indicated
CC in the FT of N81533. The first 18 N-terminal AAs have a strong
CC hydrophobic character and presumably comprise a cleavable signal
CC peptide. The extreme carboxy-terminal part comprises a region of 20
CC hydrophobic AAs, which presumably serves as a transmembrane anchor.
CC The FIVP peplomeric protein has 35 potential glycosylation sites,
CC of which 22 are in the N-terminal part (pos. 1-790) which corresponds
CC to the S-part of the IBV E2 (see P81183). N.B. IBV = infectious
CC bronchitis virus. "X" in the AA sequence denotes the translation
CC of a stop codon.
SQ Sequence 1594 AA;

```

alignment_scores:

```

Quality: 100.00 Length: 581
Ratio: 0.376 Gaps: 27
Percent Similarity: 45.783 Percent Identity: 18.072

```

alignment_block:

US-09-323-427-1 x P81184 ..

Align seg 1/1 to: P81184 from: 1 to: 1594

```

194 ACTACACTATTGATTCATTCGTTATTCGATTCGCGTTGACAAAGTGTCGA 243
51 SerThrIleMetAlaLeuSerHis.....AsnSerValIle 62
244 AGGTGAGCCAGAAATGAATGATGAGCAACTTCATTAACATCAATTTTA 293
62 uAspThrProLysPheLysCysArgCysAsnAsnAlaLeuIle ValAsn 78
294 ATACACGATATGATTCGAGAGCATGTTATGTGAAGGTCTTATGAT 343
79 LeuLysGlnLysGlnLeuAsnGlnMetVal.....ValGlyLeuIle 92
344 CAGAAGGTGGCGTAATGATGAGGTGAGCGTCAAGTTCGCCGAATTC 393
92 uArgLysGlyLysLeuLeuIleArgAsnAsnGlyLysLeuLeuAsnPheG 109
394 ACTTCATTTGATTCATTCGAATGTTGCGGTACACAGCATCTCGAATCAC 443
109 lYAsnHisLeu.....ValAsn 114

```

```

444 GGGGTTTTTGTAAACAACAAGTGTTCATTTCGTTTCATTCAT... 488
115 ValPro**LeuCysSer**LeuAlaSerCysCysTyValHisThr 131
489 ...TATTGTTACCAAAAGTTGATGTCATATCAAGTACAAATGCTTTA 534
131 rGlnPhe**ValGlnGlnIleMetAsnAlaTyrLysLeuThr**HisA 148
535 CATGCAAGCTGATTAACAAGTTA.....GTGCACAGATTGAGGATGTG 578
148 snrPrpEuaLamEtyThrLeuSerGluIlePheCysLeuValThrLeu 164
579 AAATGACA.....CTGCTTTCAACCAATATGTCGCCATG... 617
165 LysLysLysGluVal**LeuLeuValValIleThrLeuGlnAlaGlyCyl 181
618 .....CAAGTATGCCGTTATGCAATTT..... 638
181 yThrThrAlaLeuGlnGlnLeuGlnLeuLeuProPheSerIleLeuIle 198
639 .....TGG.....ATGGTG 647
198 letYmePrpPheIleLeuLeuTyrPlyProTyrPlyIleAlaLeuVal 214
648 GACCAACCGGTCAACCAAGTTCAATTGCTATCATGTGTACAGCAATTAAT 697
215 MetHisValValAsnHisIleTyrThrPheMetCysMetValSerLeuLeuVa 231
698 CAT..... 701
231 lleuLeuTyrTrpArgLeuIleGlyMetMetCysAsnLysClyProPhe**A 248
702 ...AATGACATGGGATTCTGAACACCGTTGATCTTCTGCGCG... 743
248 smetCly**CysAla**LeuLysIleAlaIleLeuThrMetAsnGln 264
743 ..... 743
265 SerProPheThrSerClyIleProHisValArgValLeuThrGluLysPhe 281
744 .....TTGTCATATCCGCTGTTTGCATGATGTAACGGGATGACTGT 786
281 eleuSerLeuSerTyrProArgThrMetGlnGlnLysSerMetValLeuS 298
787 GGAATTTCTAAATGCTGATGATGATGCTCTTGATATAATATTGCTAAATA 836
298 erClyMetThrLeuLeuGlnLeuIleLeuValValLeu..... 312
837 ATTGGAATATCCACACAGATTAAAGCGTGGCGCAAGCTCACAGTATAC 886
313 .....IleThr**Thr 316
887 AAATATGCGGATCGAT.....CACAGCTTTTCTATCATGCGCAGAT 927
316 rSerIleLeuIleGlyLeuThrMetSerHisPheCysIleHisAlaGlnA 333
928 CAGTATATACA.....TTAAAGAACCAAAATAGCGAATGTGTTGCAC 968
333 leuLeuLeuProGlyAsnThrValLeuHisMetLeuThrLysValPhe... 348
969 CACAATGTTCAACAACACACAGATTCGGAGCGTGTAAACACAGCTGTGCC 1018
349 .....LeuThrSerLeuIleThrSer**IleThrProMetVal**Ly 363
1019 GCAGCAAAACCTGCTGACGCTGCCACACTTCTGTTACTGCAGAAAAGAT 1067
363 sPrpMetAsnTyrValLysIleMetAsnIleAlaLeuAlaMetLeuProM 380
1068 .....CTCGAAGAACCGGAGAATATCATGATGTACAGACATGATATCA 1109
380 eTyrLeuLeuLeuArgHisGlnValValThrTyrLeuMetAsnLeuValLeu 396

```

```

1110 ACR.....COSTCAATTTGCGATGATATACACT 1141
1111 |||.....|||.....|||
397 ThrIleGlySerCysLeuGlnIleValProLeuLeuIleValAlaGlyLe 413
1142 TTGSCAGTGAATTATTCGTCACCGTCGACTTTCTGCAACATATAGGCAACC 1191
413 U***GlnIleAsnHisIyr**LeuIleAlaCysGlyGlnCysProVal 430
1192 TGTATACCTTGCTGCGATACAA...ATGCAATCTGACATGTCACCAATTG 1238
430 eu.....Val**GlnHisIlysAsnPrheValLeuIysValHisSerLeu 444
1239 CGTTCT.....CAATGTTATGGGTTTAAAGCATTCGATTCGTCGCC 1282
|||||.....|||||.....|||
445 AlaAsnValMetValCysLeu***IleThrGlnTrpMetLeuAspSe 461
1283 GTCATATTATTCACCTTTCGGTTTA...AATTGCGCAAAATGACAGAGCAATA 1329
:::||||::: ::::: ::::: |||::: |||::: :::::
461 rThreIleIleSerLeuGlnMetIyrAsnIleValItrValIleuGlnIyr 478
1330 AAAATAAATGTTAGATCAATCGAAGCAATAATATAAACCTGCATATATTC 1379
::: ::::: ::::: ::::: |||
478 henis**IleGlnGlnValIValSerPrheLeuIysPrheHisValIleVal 494
1380 GTTGTCTGTTATGCATCCTTCTATACCTAATTGTTAGCTACAAATATATA 1429
::: :::::
495 ThrGln***.....::: 497
1430 GATATAGCAATATATTACTGTATATACATATAGATATTTTCATCAAA 1479
::: ::::: ::::: |||
498 .....ValSerLeuValIleuThrValIleMetValIys 508
1480 CTCTCTGATGCGTTTATAGCTTTCGAAAAGCTTATTCATTAATTCAGTA 1529
::: ||| ||| ::::: ||| ::::: |||
508 erArgSer...Ala**LeuThrAspHisAspThrValIleMetIyr..... 521
1530 ATCTTTTATATGCATACCTATATTTGAATGTTTCATGTCATAGGCCATGA... 1576
||| ||| ::::: |||::: |||
522 ..PheThrMetAlaGlnLeuLeuAsnIle**GlnHisTyrHisProVal 537
1577 .....AAAGTTGCG.....TTGTATTATTCATCA 1601
537 I**ArgIysLeuLeuIleuValSerGlyAlaIlePheIleuMetValT 554
1602 TTATCAACTGTCCTATTATTATTATCAACTTTATTCATTG 1642
||||:||||:||||:||||:||||: |||
554 hIleSerLeuAlaHisPheLeuLeuValValIyrLeuLeu 567

seq_name: A_Geneseq_36:R87953

seq_documentation_block:
ID R87953 standard; Protein: 332 AA.
AC R87953;
DT 20-MAR-1996 (first entry)
DE Bovine neurotrophic biglycan.
KW Biglycan; proteoglycan; chondroitin sulphate; neuron protection;
  neurotrophic; central nervous system; CNS; memory loss; dementia;
  learning.
OS Bos taurus.
FH key Location/Qualifiers
FT region 7..23
FT /label= Hypervariable_region
PD WO9530432-AI.
PD 16-NOV-1995.
PF 09-MAY-1994; E01479.
PR 09-MAY-1994; WO-E01479.
PA (BOE ) BOEHRINGER MANNHEIM GMBH.
PI Haenoehehl R, Huston J, Jungmans U, Kappler J, Koops A;
  Mueller HW;
  WPI: 95-403938/51.
PT proteoglycan cpds., partic. chondroitin sulphate proteoglycans)
  for maintain structural and function of the CNS and attenuating
  memory deficit(s) in the elderly and patients with dementia
  PT

```

PS Claim 3; Fig 8; 60pp; English.
 CC Bovine biglycan (R87953) is a chondroitin sulphate proteoglycan with
 CC neurotrophic activity for brain neurons. It can be used to enhance
 CC the survival and maintain the structure and function of CNS neurons
 CC during normal ageing as well as after pathological and/or traumatic
 CC nervous system damage. It can also be used to restore function
 CC following nervous system lesions and degenerative diseases, and to
 CC improve learning efficiency and memory in the elderly and in patients
 CC with dementia.
 SQ Sequence 332 AA;

alignment_scores:
 Quality: 96.50 Length: 299
 Ratio: 0.798 Gaps: 13
 Percent Similarity: 40.468 Percent Identity: 20.401

alignment_block:
 US-09-323-427-1 x R87953 ..

Align seg 1/1 to: R87953 from: 1 to: 332

```

1005 AACACAGGTGTCGCCGACCAAACTGTCAGCTGCCGCACTTGCTTAA 1054
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
50 LysGluIleSerProAspThrThrLeuLeuAspLeuGlnAsnAsnAsp11 66
1055 CTCA.....AGAAAAGAT..... 1067
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
66 eSerGluLeuArgLysAspAspPheLysGlyLeuGlnHisLeuTyrAlaL 83
1068 .....CTGCAGAACCGGAGAAATATCATTTAGTACGAACTGATACACA 1112
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
83 euValLeuValAsnAsnLysIleSerLysIleHisGluLysAlaPheSer 99
1113 CCCTTGAATAGTCGATGATTAATGCAAGCTTGGCAGTTGATTACCTCAC 1162
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100 ProLeuArg.....LysLeuGlnLysLeuTyrIleSerLys 111
1163 CGTGCACCTTCTGCACATATATGACACACTGTATATACCTGTCAGTACA 1212
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
111 sAsnHisLeuValGluIleProProAsn..... 120
1213 AATGGAATCTGCATCTCACCATTTGGCTTTCATGTTTATGCGTTTAA 1262
120 ..... 120
1263 GCATTCATTTGATTCCTGCCGTCATATTATACCATTTGTTAAATTCGT 1312
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 .....LeuProSerSerLeu.....Va 126
1313 CCAAAATCAGAGGCAATMAAATATATGTTAGA.....ATCA 1347
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
126 lGluLeuArgLysIleHisAspAsnArgIleArgLysValProLysGlyValP 143
1348 TCACAACCATTA..... 1358
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
143 heSerGlyLeuArgAsnMetAsnGluIleGluMetGlyGlyAsnProLeu 159
1359 .....ATAAACTGCCATA 1372
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
160 GluAsnSerGlyPheGluProGlyAlaPheAspGlyLeuLysLeuAsnTyr 176
1373 TATATTCGTTTCTTTATCATCTTCTATATACTAATTTTACSTACAA 1422
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
176 rLeuArgLysSerGluAlaLys.....LeuThrG 186
1423 ATATATAGATGTAGGAATATATCTATATACATATAGATATTTTC 1472
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
186 lTyrLe.....ProLysAspLeuProGluThrLeuAsnGluLeuHis 199
1473 ATCAAAACTCTCTCTATCGCTTTATGAGCTTCAAAAAGCTTATTCATTA 1522
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
200 LeuAspHisAsnLysIleGlnAlaIleGluLeuGlnAspLeuLeuArgTyr 216

```

```

1523 TTCAGTAACTTTTATATGCACTACTATGTAAATGTTTCATCATTAAGCC 1572
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
216 rSerLysLeuTyrArgLeu.....GlyLeuGlyH 226
1573 ATGAATACCTTTCGTTTGTATATCA...TCATATACACTTGTCT... 1616
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
226 lAsnGlnIleArgMetIleGluAsnGlySerLeuSerPheLeuProThr 242
1616 ..... 1616
243 LeuArgGluLeuHisLeuAspAsnAsnLysLeuSerArgValProAlaG1 259
1617 .....ATTATTTCTAACAGTTTATTCATTTGTCATATA 1650
259 yLeuProAspLeuLysLeuLeuGlnValValTyrLeuHisThrAsnAsn1 276
1651 TCACAATTTATACCTTG.....TATTCGCCAATTTTATGGGCATC 1691
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
276 lThrLysValGlyValAsnAspPheCysProValGlyPheGlyVal 291

```

seq_name: A_Geneseq_36:R87951

seq_documentation_block:

```

ID R87951 standard; Protein; 369 AA.
AC R87951:
DE 20-MAR-1996 (first entry)
DE Rat neurotrophic biglycan.
DE Biglycan; proteoglycan; chondroitin sulphate; neuron protection;
KW neurotrophic; central nervous system; CNS; memory loss; dementia;
KW learning.
OS Rattus sp.
FH Key location/Qualifiers
FT peptide 1..37
FT /label= Sig-peptide
FT region 44..60
FT /label= Hypervariable_region
PN WO9530432-A1.
PD 16-NOV-1995.
PR 09-MAY-1994; E01479.
PR 09-MAY-1994; WO-E01479.
PA (BOEF) BOEHRINGER MANNHEIM GMBH.
PI Hasenoehtl R, Huston J, Jungmans U, Kappler J, Koops A;
PI Mueller HW;
DR WPI: 95-403938/51.
DR N-PSDB: T08768.
DR Proteoglycan cpts., partic. chondroitin sulphate proteoglycan(s)
PT for maintain structural and function of the CNS and attenuating
PT memory deficit(s) in the elderly and patients with dementia
PS Claim 1; Page 44-45; 60pp; English.
CC Rat biglycan (R87951) is a chondroitin sulphate proteoglycan with
CC neurotrophic activity for brain neurons. Recombinant biglycan,
CC obtd. by expression of encoding cDNA (T08768) in eukaryotic host
CC cells, can be used to enhance the survival and maintain the structure
CC and function of CNS neurons during normal ageing as well as after
CC pathological and/or traumatic nervous system damage. It can also
CC be used to restore function following nervous system lesions and
CC degenerative diseases, and to improve learning efficiency and memory
CC in the elderly and in patients with dementia.
SQ Sequence 369 AA;

```

alignment_scores:
 Quality: 95.50 Length: 299
 Ratio: 0.789 Gaps: 13
 Percent Similarity: 40.468 Percent Identity: 20.401

alignment_block:

US-09-323-427-1 x R87951 ..

Align seg 1/1 to: R87951 from: 1 to: 369

1005 AACACAGGTGTCGCCGACCAAACTGTCAGCTGCCGCACTTGCTTAA 1054

```

|||||: |||: ::|||:|||||:
87 LysGluIleSerProAspThrThrIleuAspLeuGlnAsnAsnAsp11 103
1055 CTCGAGGAT.....AGAAAGAT..... 1067
103 eSerIleuArgIlyAspAspPheIySGIleuGlnHisLeuYrAlAl 120
1068 .....CTGCAGAACCGGAGAAATATCATGATGTAGCAACTGATATCAACA 1112
120 euValleuValAsnAsnIlySleSerIySIlleHisGluIySAlaPheSer 136
1113 CCCTTGAAATTAGCGATGATTAATCAAGCTTGGCCAGTTGATTTACGTCAC 1162
137 ProLeuArg.....LysLeuGlnIlySLeuYrIleSerIy 148
1163 CGTGCACTTCTGCACATTAATGAGACACCTGTAATACTGTCGCACTACA 1212
148 sAsnHisLeuValGluIleProAsn..... 157
1213 AATGGAATCTGCATGTCACACATTTGGCTTCAATGTTATGGGTTTAA 1262
157 ..... 157
1263 GCATTCGATGATTCGTCGCCGTCATTAATACATTTCCGTTAAATTCGT 1312
158 .....LeuProSerSerLeu.....Va 163
1313 CCAATTCAGAACGATCAATAAATAATGTTAGA.....ATCA 1347
163 lGluLeuArgIleHisAspAsnArgIleArgIyValProIySGIyAlP 180
1348 TCGAAGCAATA..... 1358
180 heSerGIyLeuArgAsnMeIAsnCysIlleGluMeGIyGIyAsnProLeu 196
1359 .....ATRAAACTGCCATA 1372
197 GluAsnSerGIyPheGluProGIyAlaPheAspGIyLeuIySLeuAsnIy 213
1373 TAAATTCGTTCTCTATATCATCCTTCTAATACTAATTTAGTACACA 1422
213 rLeuArgIleSerGIuAlAlAs.....LeuThrG 223
1423 AATATATAGATGTAGCAATAATAATTAATCAATAAGATATTTTC 1472
223 lYIle.....ProIySAspLeuProGIuThrLeuAsnGIuLeuHis 236
1473 ATCAAAATCTTCTATGCGTTTATAGCTTCTGAAAGCTTATTCATTA 1522
237 LeuAspRHisAsnIlySIlleGlnAlAlleGIuLeuGIuAspLeuLeuArgIy 253
1523 TTCAGTAATCTTTATATGCACTATATGTAATGTTTCATCATTAAGGCC 1572
253 rSerIySLeuYrArgLeu.....GlyLeuGIyN 263
1573 ATGAATAGTTTGGTTGTATATATCA...TCATTAATCAACTGTCTCT... 1616
263 lAsnGIuIleArgMeTIleGIuAsnGIySerLeuSerPheLeuProThr 279
1616 ..... 1616
280 LeuArgGIuLeuHisLeuAspAsnAsnIlySLeuSerArgValProAlAl 296
1617 .....ATTTTCTTAACAGTTTATCATTCATTTGATTAATA 1650
296 yLeuProAspLeuIySLeuLeuGlnValIyTrLeuHisSerAsnAsnI 313
1651 TCACAAATTTATACCTTG.....TATGGCCAAATTTTATGGGCATC 1691
313 lEThrIySValGIyIleAsnAspPheCysProMeCglYPheGIyAl 328
seq_name: A_Geneseq_36:R87952

```

```

seq_documentation_block:
ID R87952 standard; Protein: 369 AA.
AC R87952;
DE 20-MAR-1996 (first entry)
DE Human neurotrophic biglycan.
KW Biglycan; proteoglycan; chondroitin sulphate; neuron protection;
KW neurotrophic; central nervous system; CNS; memory loss; dementia;
OS Homo sapiens.
FT Key location/Qualifiers
FT peptide 1..37
FT /label= Sig_peptide
FT region 44..60
FT /label= Hypervariable_region
FN WO9530432-A1.
PD 16-NOV-1995.
PE 09-MAY-1994; E01479.
PR 09-MAY-1994; WO-E01479.
PA (BOE) BOEHRINGER MANNHEIM GMBH.
PI Hasenoechl R, Huston J, Jungmans U, Kappler J, Koops A;
PI Mueller HW;
DR WPI: 95-40398/51.
PT Proteoglycan cpts., partic. chondroitin sulphate proteoglycan(s)
PT for maintain structural and function of the CNS and attenuating
PT memory deficit(s) in the elderly and patients with dementia
PS Claim 3; Fig 8; 60pp; English.
CC Human biglycan (R87952) is a chondroitin sulphate proteoglycan with
CC neurotrophic activity for brain neurons. It can be used to enhance
CC the survival and maintain the structure and function of CNS neurons
CC during normal ageing as well as after pathological and/or traumatic
CC nervous system damage. It can also be used to restore function
CC following nervous system lesions and degenerative diseases, and to
CC improve learning efficiency and memory in the elderly and in patients
CC with dementia.
SQ Sequence 369 AA;

```

```

alignment_scores:
Quality: 94.50 Length: 299
Ratio: 0.781 Gaps: 13
Percent Similarity: 40.468 Percent Identity: 20.401

```

```

alignment_block:
US-09-323-427-1 x R87952 ..

```

```

Align seg 1/1 to: R87952 from: 1 to: 369

```

```

1005 AACAGGTGTCGCCGACGCAAACTGTCAGCTGCCGCACTTGTTTA 1054
|||||: |||: ::|||:|||||:
87 LysGluIleSerProAspThrThrIleuAspLeuGlnAsnAsnAsp11 103
1055 CTCGAGGAT.....AGAAAGAT..... 1067
103 eSerIleuArgIlyAspAspPheIySGIleuGlnHisLeuYrAlAl 120
1068 .....CTGCAGAACCGGAGAAATATCATGATGTAGCAACTGATATCAACA 1112
120 euValleuValAsnAsnIlySleSerIySIlleHisGluIySAlaPheSer 136
1113 CCCTTGAAATTAGCGATGATTAATCAAGCTTGGCCAGTTGATTTACGTCAC 1162
137 ProLeuArg.....LysLeuGlnIlySLeuYrIleSerIy 148
1163 CGTGCACTTCTGCACATTAATGAGACACCTGTAATACTGTCGCACTACA 1212
148 sAsnHisLeuValGluIleProAsn..... 157
1213 AATGGAATCTGCATGTCACACATTTGGCTTCAATGTTATGGGTTTAA 1262
157 ..... 157
1263 GCATTCGATGATTCGTCGCCGTCATTAATACATTTCCGTTAAATTCGT 1312
|||||: |||:

```

```

158 .....LeuProSerSerLeu.....Va 163
1313 CCAAAATCAGAGGCATATAAATAATTGTTAGA.....ATCA 1347
      ||||| ||| |||:::|||||
163 IGIleuArgIleHisAspAsnArgIIeArgLysValProlYsglyAlaP 180
1348 TCGAGCAATAA..... 1358
180 heserGlyLeuArgAsnMetAsnCysIleGluMetGlyLysProLeu 196
1359 .....ATAAACGTGCCATA 1372
      ::||| |||
197 GluasnsereGlyPheGluProGlyAlaPheAspGlyLeuLysLeuAsnTy 213
1373 TAATATTGGTTCCTTCATCATCCCTCTTAATACATAATTATTCAGTAA 1422
      |||||:::|||||
213 rLeuArgIleSereGluAlaLys.....LeuInrG 223
1423 ATATATGATGATGTAGCAATAATTACTGTATATACAAATAGATATTTC 1472
      :||| ::::: |||||:::
223 Lyle.....ProlYsgAspleuProGluTrpLeuAsnGluLeu 236
1473 ATCAAACCTTCCTTCATCGCTTTTATAGCTTCGAAAAGCTTATTCATA 1522
      ::::: ||||| ||| |||:::
237 LeuAspHisAsnLysIleGlnAlaIleGluLeuGluAspLeuLeuArgTy 253
1523 TTCAAGTATCTTTTATATGACACTACTATTGTGAATGGTTTCAATATGACC 1572
      |||| ::::: ::|
253 rSerLysLeuTyArgLeu.....GlyLeuGlyH 263
1573 ATGAATAGTTTCGTTTGTTATATCA...TCATATGCACTTGCT... 1616
      ||::: |||||::: ::| |||||
263 IsAsnGlnIleArgMetIleGluAsnGlySerLeuSerPheLeuProThr 279
1616 ..... 1616
280 LeuArgGluLeuHisLeuAspAsnAspLysLeuSerArgValProAlaGl 296
1617 .....ATTTTATTTCTAACGTTTATCATTTTGTGATATA 1650
      ::||| :::::
296 yLeuProAspLeuLysLeuGlnValTyLeuHisSerAsnAsnI 313
1651 TCACAATTTATACCTTG.....TATTGCCCAATTTTTATGGGATC 1691
      |||||::: ::|||
313 lethrIvalGlyValAsnAspPheCysProMetGlyPheGlyVal 328
seq_name: A_Geneseq_36:P94856
seq_documentation_block:
ID P94856 standard; protein: 693 AA.
AC P94856:
DT 27-JUN-1990 (first entry)
DE Expression plasmid pUCPICIT799 BglII gene product.
KW alpha-amidating; PAK799; alpha amide; ds.
OS Synthetic.
PN EP-299790-A.
PD 18-JAN-1989.
PF 15-JUL-1988; 306508.
PR 17-JUL-1987; JP-177184.
PR 05-DEC-1987; JP-306867.
PA (SUNR) Suntory Ltd.
PI Ohshuye K, Kitano K, Tanaka S, Matsuo H, Mizuno K;
   WPT: 89-017279/03.
DR N-BSD: N94527.
RT Recombinant C terminal alpha amidating enzymes of Xenopus laevis -
   PT and their precursors deoxyribonucleic acid encoding sequences.
PS Disclosure; ?p; English.
CC The sequence encodes a derivative of the mature C-terminal alpha-
   CC amidating enzyme from plasmid PKA799.
CC The plasmid was screened from an E.coli library using plasmid PKA457
   CC to screen a larger library.
CC Although PKA799 is similar to PKA457 at the N-terminus, it has an area of
   CC hydrophobic elements suggesting a membrane function.
```

[illegible]

ID P94854 standard; protein; 875 AA.
 AC P94854;
 DT 27-JUN-1990 (first entry)
 DE C-terminal prepro-C-terminal alpha-amidating enzyme of pX4799.
 KM alpha-amidating; pX4799; alpha amide; ds.
 OS Synthetic.
 PN EP-299790-A.
 PD 18-JAN-1989.
 PE 15-JUL-1988; 306508.
 PR 17-JUL-1987; JP-177184.
 PR 05-DEC-1987; JP-306867.
 PA (SUNR) Suntory Ltd.
 PI Ohsuye K, Kitano K, Tanaka S, Matsuo H, Mizuno K;
 DR WPI: 89-017279/03.
 N-PSDB: N90791.
 PT Recombinant C terminal alpha amidating enzymes of *Xenopus laevis* -
 and their precursors deoxyribonucleic acid encoding sequences.
 PS Disclosures, 7pp. English.
 CC Plasmid pX4799 contains a sequence derived from *Xenopus laevis*.
 CC The plasmid was screened from an E.coli library using plasmid pX457
 CC to screen a larger library.
 CC Although the gene product is similar to that of pX457 at the N-terminus,
 CC it has an area of hydrophobic elements suggesting a membrane function.
 CC See also N93060.
 SQ Sequence 875 AA;

alignment_scores:
 Quality: 94.50 Length: 232
 Ratio: 0.945 Gaps: 12
 Percent Similarity: 43.103 Percent Identity: 21.552

alignment_block:

US-09-323-427-1 x P94854 ..

Align seg 1/1 to: P94854 from: 1 to: 875

```

263 TGTGACCACTTCAATACATCAATTTTAATACACGTAATGCAATGCA 312
    ||| |||||:|||||:||||| ||| ||| |||
530 CysGlnProThrAspValAlaValAspProIleThrGlyAsnPhenPheVa 546
313 AGGACATGTTTATGTGAAGTCTTTATGATCAAGAGGTGCCGTAATG 362
    :|||
546 lAlaAspGlyTyr..... 550
363 ATGAGGTGACGTCAGTGGCGGAATTCACCTTCATTTGATTCATGAC 412
551 ..... 551
413 AATGTTGCCGTCACAGATCTCTGTAATCCAGCTGATTTTGTGA..... 457
    ||| :||| |||:|||||:|||||:|||||:|||||:|||||:
552 Asn...SerArgIleMetGlnPheSerProAsnGlyMetPheIleMetG 567
457 ..... 457
567 nTrpGlyGlnGluThrSerSerAsnLeuProArgProGlyGlnPheArgI 584
458 .....ACAAACAAGTGTTCATTTGCTTCATTCATTCATTTGTTTACC 499
584 lerpIHisSerLeuThrMetIleSerAspGlnGlyGlnLeuGlyValAla 600
500 AAAGTTGATCGTCATACGATACATGCTTTTACATGGAAGCTGATTA 549
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
601 AspArgGluAsnGly...ArgIleGlnCysPheHisAlaLeuThrGly 616
550 AACACTTATGACAGATGAGTATCTGAATC...ACAACCTGCTTTTC 596
    : |||:|||||:|||||:|||||:|||||:|||||:
616 uPheValIys...GlnIleLeuHisGlnIlePheGluArgGluValPhe 632
597 AAACCAAAATGTTCCGATGCCAGTATGCCGTTATGAATTTTGATGCT 646
    :|||:|||||:|||||:|||||:|||||:|||||:
632 lAlaSerTyrAlaPro.....Gly 638

```

```

647 GGACCAACCGGTCAACCACTTCATATTGCTATCATTTGTCACCCAGTTTA 696
    ||| ||| :|||:|||||:||||| ||| ||| |||
639 Gly.....ValLeuYrAlaValAsnGlyLysProTyr 650
697 TCATTAATGACATGCGATTCGAAACCGTTCATCTTCGCGGGGTG 746
    | ||||| |||:||||| |||
650 rGly.....AspSerThrProValGlnGlyPhe..... 659
747 TCCATTCCTGCTTTGTCATGATGTAACGTCGATACGTAATTCGA 796
    :|||:|||||:|||||:|||||:|||||:|||||:
660 .....MetLeuAsnPhenSerAsnGlyAsp..... 667
797 AATGCTGATGATGTCCTTCGATAAATATTG.....CTAAATATT 840
668 .....IleLeuAspThrPheIleProAlaArgLysAsnPh 679
841 GGAATATCCACAGATTTTAATAGCTGGCCAAACCTGACGTAAC 886
    :||| ||| |||:|||||:|||||:|||||:|||||:
679 eGlnMetProHisAspIleAlaGlyAspAspGlyThrValTyr 694

```

seq_name: A_Geneseq_36:R73053

seq_documentation_block:

ID R73053 standard; Protein; 875 AA.
 AC R73053;
 DT 06-NOV-1995 (first entry)
 DE Peptidyl C-terminal alpha-amidating enzyme.
 KW Peptidyl C-terminal alpha-amidating enzyme; AE: cell culture;
 KW trichostatin, CHO.
 OS Not specified.
 FH Key Location/Qualifiers
 FT peptide 1..39
 FT EP-649900-A.
 PD 26-APR-1995.
 PE 07-SEP-1994; 306587.
 PR 08-SEP-1993; JP-257881.
 PA (FURU//) FURUKAWA K.
 PA (SUNR) SUNTORY LTD.
 PI Furukawa K, Ohsuye K, Sugimura K;
 DR WPI: 95-156754/21.
 DR N-PSDB: Q87970.
 PT Increasing protein prodn. from cultured animal cells - by adding
 PT a trichostatin to the medium, effective at low concn. and not
 PT injurious to host cells
 PT Disclosures: Page 10-15; 19pp. English.
 CC 3mu-1S cells (i.e. CHO cells expressing a peptidyl C-terminal alpha-
 CC amidase enzyme (AE)) were suspended in F-12 medium to which aliquots
 CC of trichostatin were added. Cells were cultured for 3 days at 37 deg
 CC and then assayed for AE. Without trichostatin, AE productivity was
 CC 708 U/ml. At 10, 50, 100 and 200 nM trichostatin, AE productivities
 CC were respectively 866, 1897, 1894 and 3359 U/ml.
 SQ Sequence 875 AA;

alignment_scores:
 Quality: 94.50 Length: 232
 Ratio: 0.945 Gaps: 12
 Percent Similarity: 43.103 Percent Identity: 21.552

alignment_block:

US-09-323-427-1 x R73053 ..

Align seg 1/1 to: R73053 from: 1 to: 875

```

263 TGTGACCACTTCAATACATCAATTTTAATACACGTAATGCAATGCA 312
    ||| |||||:|||||:||||| ||| ||| |||
530 CysGlnProThrAspValAlaValAspProIleThrGlyAsnPhenPheVa 546
313 AGGACATGTTTATGTGAAGTCTTTATGATCAAGAGGTGCCGTAATG 362
    :|||
546 lAlaAspGlyTyr..... 550

```

```

363 ATGAAAGTGCACGTCAAGTTGCCGGAATTTCACTTCATTGATGATGC 412
551 .....Cys 551
413 AATGTTGGCGTACACGATCTGAAATCCAGCTGATTTTGTAA..... 457
552 Asn...SerArgIleMetGlnPheSerProAsnGlyMetPheIleMetG 557
457 ..... 457
567 nTrpGlyGlnGlnThrSerSerAsnLeuProArgProGlyGlnPheArg 584
458 .....ACAAACACTGTGTTCATTTCGTTTCATCCATTATTTGTTAC 499
584 IeProHisSerLeuThrMetIleSerAspGlnGlyGlnLeuGlyValAla 600
500 AAGATTGATCGTCATATGACGATGACATGCTTTTACATGGAACCTGATA 549
601 AspArgGlnAsnGly...ArgIleGlnCysPheHisAlaLysThrGlyG 616
550 AACAGTTAGTGCACAGATTGAGTATCTGAAATC...ACAACTGCTTTTC 596
616 unPheValys...GlnIleLysHisGlnGlnPheGlyArgGlyValAla 632
597 AAACGTAATGTGCCGATGCCAGTATGCCGTTATGAAATTTTGATGCT 646
632 IValSerIleTyrAlaPro.....Gly 638
647 GGACCAACCGGTCAACCAAGTTCAATTGTCATGATGTCAGCCAGTTTA 656
639 Gly.....ValLeuTyrAlaValAlaAsnGlyLysProTyrGly 650
697 TCATAATGACATGCGAATTCGAAACCGTTGATACTTCTCGCGCGATTG 746
650 rGly.....AspSerThrProValGlnGlyPhe..... 659
747 TCCATTCTGCTTTGTCGATGATGATGACGCGTATCTGGAATTCGA 796
660 .....MetLeuAsnPheSerAsnGlyLysP..... 667
797 AATGCTGATGATGCTGCTCTGTGATAAATATTG.....CTAATAATTT 840
668 .....IleLeuAspThrPheIleProAlaArgLysAsnPr 679
841 GGAATATCCAAACGATTTAATGCGTGGCCAAAGCTCAGCTATAC 886
679 eGluMetProHisAspIleAlaIleAlaGlyAspArgLysThrValTyr 694

seq_name: A_Geneseq_36:W13825
seq_documentation_block:
ID W13825 standard; Protein; 1226 AA.
AC W13825:
DE 04-JUN-1997 (first entry)
DE Yeast transcription regulatory factor SRB8.
DE Transcription regulatory factor; suppressor of RNA polymerase B;
DE SRB8; RNA polymerase II; holoenzyme; SWI/SNF.
OS Saccharomyces cerevisiae.
PN M09708301-A1.
PD 06-MAR-1997.
PE 28-AUG-1996; 014192.
PR 31-AUG-1995; US-521872.
PR 11-OCT-1995; US-540804.
PR 26-JAN-1996; US-590399.
PA (NHED ) WHITEHEAD INST BIOMEDICAL RES.
PI Chao DM, Koleske AJ, Thompson CM, Young RA.
PI WPI: 97-179258/16.
DR N-PSDB; T59908.
PT Purified RNA polymerase II holo:enzyme - comprises RNA polymerase II
PT and one or more regulatory proteins, pref. suppressor of RNA
PT polymerase B proteins or SWI/SNF proteins
PS Claim 11; Fig 10a-b; 154pp; English.
CS Novel yeast SRB (suppressor of RNA polymerase B) proteins SRB4, SRB5,

```

```

CC SRB6, SRB7, SRB8, SRB9, SRB10 and SRB11 (W13821-28) are transcription
CC regulatory factors that act as positive and negative regulators of
CC RNA polymerase II activity, and are components of the RNA polymerase
CC II holoenzyme. They were identified using methods designed to
CC identify transcription factors involved in RNA polymerase II
CC C-terminal domain (CTD) function. SRB8 and SRB9 appear to repress
CC CTD activity. Genomic clones (T59904-11) for the SRBs have been obd.
CC SRBs can be used to treat diseases resulting from alteration or
CC deletion of the SRB gene, pref. by gene transfer technology. They
CC can also be used in in vitro transcription of DNA and to identify
CC cps. that modify gene transcription.
CC Sequence 1226 AA:

alignment_scores:
Quality: 91.50 Length: 489
Ratio: 0.393 Gaps: 24
Percent Similarity: 47.648 Percent Identity: 19.632

alignment_block:
US-09-323-427-1 x W13825
Align seg 1/1 to: W13825 from: 1 to: 1226

513 CATATGAGTACAAATGCTTTTACATGGAAGCTGAAACAGTTA..... 557
.....:|||||:|||||
279 TyrValAspGlnAsnProLeuThrMetHisLysIleIleGlnLeuIle 295
558 .....GTGCACAGATTGAGC.....TATCGAATTCAGAC 588
.....:|||||:|||||
295 uTrpSerIleHisProSerArgGlnPheAspHisTyrGluSerAsnGlnL 312
.....:|||||:|||||
589 TCGTTTTCAACTCAAAATTTGTCGCCAGTATGCCGTTATGAAATTT 638
|||:|||||:|||||
312 euValAlaLysLeuLeuLeuLeuArgIleAsn..... 322
639 TGGAGTGGACCAACCGGTCAACCAAGTTCAAT.....TTGCT 676
.....:|||||:|||||
323 ...SerThrAspGlnAsnProLeuHisGlnPheGlnIleGluAspAlaIle 338
.....:|||||:|||||
677 ATCATGTGTCACCCAGTTTATCATCAATGACATGCGATTCGAAACGCT 726
|||||:|||||:|||||
338 pSerLeuValPheGlnLeuAlaLysAsn..... 347
727 TATATCTTCTGCGCGTTGTCATCTCTGTCGATGATGATGATGATGATG 776
|||:|||||:|||||
348 .....PheSerAlaGlnLysArgValAlaSer 356
777 GTGATACGTGGAATTTCTAAATGCTGATGATGATGCTCTTGAATAATAT 826
.....:|||||:|||||
357 TyrMetMetProSerLeuTyrArgLeuLeuAsnIleLeu..... 369
827 TTGCTTAATAATTTTGAATATCCAAACAGATTTAATGCGTGGCCAGAAC 876
|||:|||||:|||||
370 .....IleThrTyrGlyIleIleLysValProThrTyrIleArgLysL 384
877 TCACGTAT.....ACAAATTCGCGATCAT 902
|||:|||||:|||||
384 euIleSerSerGlyLeuLeuTyrLeuGlnAspSerAsnAspLysPheVal 400
903 CAC...AGCTTTTATCATCAATCCAGATCAGTATTACCTTAAGAAGCA 949
|||:|||||:|||||
401 HisValGlnLeuLeuIleAsnLeuLysIleSerProLeuMetLysSerG 417
950 A.....ATAGCGAATGTGTGACACACATGTTCCGAACACCA 987
|||:|||||:|||||
417 nTyrAsnMetValAlaLeuArgAsnValMetGlnTyrAspValLysPheTyrG 434
988 AG.....GATTCGAGAGCTGTTAAACAGGTGTCGCCGAGCAAAACCG 1031
|||:|||||:|||||
434 IuIlePheAsnPheAspGlnValAlaGluIleThrGluGlnIleLysMet 450
1032 CTGCAGCTGCGCAACTTCGTTACTCAAGAAAGATCTGCAGAACCGGAG 1081

```


[illegible]

seq_documentation_block:
ID R89462 standard; Protein: 2756 AA.
AC R89462;
DT 30-JAN-1997 (first entry)
DE Biosynthetic enzyme of icosapentaenoic acid synthase.
EC Icosapentaenoic acid synthase; EPA; drugs; agrochemicals;
KW foodstuffs; animal feed; lipid balance correction; antihypertensive;
KW antiinflammatory; anticancer agent.
OS *Shewanella putrefaciens*.
PN M09621735-A1.
PD 18-JUL-1996.
PF 12-JAN-1996; J00030.
PR 13-JAN-1995; JP-004299.
PT (SAGA) SAGAMI CHEM RES CENTRE.
PI Kato S, Kondo K, Yamada A, Yazawa K;
PI WPI: 96-34228/34.
DR N-PSDB; T34137.
PT Production of icosapentaenoic acid using transformed *E. coli* - uses
PT DNA coding for icosapentaenoic acid synthase derived from *Shewanella*
PT strain
PS Claim 7: Page 83-94: 145pp: English.
CC The DNA sequence (T34137) which encodes the biosynthetic enzymes of
CC icosapentaenoic acid (EPA) can be used to transform *Escherichia coli*
CC The DNA sequence allows efficient microbial production of EPA, which
CC is a raw material for drugs, agrochemicals, foods and animal
CC feedstuffs. EPA is also useful for lipid balance correction and as
CC an antihypertensive, antiinflammatory and anticancer agent.
SQ Sequence 2756 AA;

```

alignment_scores:      quality: 91.00      length: 500
                       ratio: 0.389      gaps: 23
Percent Similarity: 46.800      Percent Identity: 18.400

alignment_block:
US-09-323-427-1 x R99462 ..

Align seg 1/1   to: R99462   from: 1   to: 2756

```

35 CCGATGGGACGACCAACCAACCAACCAACCAACCAACCAACG 84
 827 P r o p h e t a l a l y s a l a l l e a s p a l a l a l y s e h r i l y s t h r s e r a l g a l 843
 85 A A T A A T A A C C C C A C T C A G T G A G G A G A G A C A G A C A A C T T A C T T T 134
 843 a l e u t y r s e r s a n l a t h r g l y g l y . . . l e u t y g l i n s e r t h r a l a l a l 859
 135 T T C T A A A A N T G A A N T T A C T A A C T T C T G A A T G A T G A T T C G T T A T T 184
 859 y s i l e l y s a l s e r p h e l y s l y s h s M e t l e u g l i n s e r v a l 872
 185 G C T T T C T C T A C T A C T A C T A T T G A C A T T G C T T A T T C G A T T C G G T T G C A A 234
 873 A r g p h e t h r s e r g l n e u g l u a l a m e t T y r s a n a s 884
 235 T G G T G C A A G S T A G C C A G A A T T G A A T G A T G A C C A C T T C A T A . . . 280
 884 p o l y A l a r g v a l P h e a l g l u p h e g l y p r o l y s a s n l l e u g 899
 281 A C A 283
 899 I n l y s l e u v a l g l n g l y t h r l e u v a l s n t h r g l u a s n g l u a l c y s t h r 915
 284 A T C A T T T T A A T A C A C G T A A T G C A T T C G A A G A G A C A T T T A T G T G A A A G G 333
 916 I l l e s e r i l l e a s n p r o l a s n p r o l y s v a l a s p e r s p e u n l n e u t y s . . 921
 334 T C T T A T A T C A A G A A G S T T G C C G A A T G A T G A T G A G G A C G C A A T T G 383
 932 G l n a l a l a m e t c l n l e u a 938
 384 C C G G A A T T T C A C T T C C A T T G A T T C A T G C A A T 415

```

1211 hhrhAlaValGIhNthAlaProIaGIaValaValaArgGIaIaIaPro 1227
1148 GTTGATTACGTCACCGTCGACTTCTGCAACATAATGACACACGTAT 1197
1128 Val.....GlnAlaIaIaIeGlnProIleAs 1236
1198 ACTTGCTGACGACAAATGGAATCTGCATGTCACCATTTGGCTTCA 1247
1236 nThSerValaIaThr.....ThrProSerAlaIaPheSer 1249
1248 TGTATTGAGTTAAGCATTCGATTGTCCTGCCGTATTAATCAAT 1297
1249 IaGIuThAlaIeSerAlaThrIaIaThrIaValGIaIaIaThrMetIeGIaVal 1265

seq_name: A_Geneseq_36:R05159

seq_documentation_block:
ID R05159 standard; protein: 368 AA.
AC R05159;
DT 09-OCT-1990 (first entry)
DE Sequence of human bone proteoglycan I (biglycan).
KW Osteoporosis; Rheumatoid arthritis; Paget's disease;
KW atherosclerosis; periodontal; human bone matrix; proteoglycan.
KW Homo sapiens.
PN U57432044-A.
PD 17-APR-1990.
PF 3-NOV-1989; 432044.
PR 3-NOV-1989; US-432044.
PA (USSH) Nat Inst of Health.
PI Termine J;
PI WPI: 90-178641/23.
DR N-PSDB: 004490.
PT Human bone matrix DNA and proteins
PT used in detection, diagnosis and treatment involving skeletal
PT and/or connective tissue disease states.
PS Disclosure; P: English.
CC Probes and Abs raised to the proteins can be used to determine
CC their levels useful in diagnosis of associated connective tissue
CC diseases states such as osteoporosis, osteo/rheumatoid arthritis
CC Paget's disease, atherosclerosis and periodontal disease.
CC Proteins may also be used to induce or block biological function
SQ Sequence 368 AA;

alignment_scores:
Quality: 90.50 Length: 299
Ratio: 0.730 Gaps: 14
Percent Similarity: 41.472 Percent Identity: 19.732

Alignment_block:
US-09-323-427-1 x R05159 ..

Align seg 1/1 to: R05159 from: 1 to: 368

1005 AATGAGTGGTGGCGGCAAAACCTGCTGACCTGGGCACTTGTTA 1054
|||||..... |||.....:|||||.....:|||||
86 LysGIuIleSerProAspThrThrIeUeAspLeuGIaIaAsnAspI 102
|||||.....:|||||.....:|||||.....:|||||
1055 CTCGA.....AGAAAGAT..... 1067
||||| |||||||
102 eSerGIuIeUaArgLysAspPheLysGIeUaGlnHisLeuTyrAla 119
||||| |||.....:|||||.....:|||||
119 euValIeUaValaAsnAspLysIleSerLysIleHisGIuLysAlaPheSer 135
|||||.....:|||||.....:|||||.....:|||||
1113 CCCTTGAAATTAGCGATGATATCAACGCTTGGCACTGATTTAGCTAC 1162
|||||.....:|||||.....:|||||.....:|||||
136 ProIeUaArgAsnVal.....GlnLysIeUaTyrIleSe 146
CGTGCACCTTTCGAACATATATGGAACAACCTGTATATCTTGCGACGTACA 1212
146 Lys..... 147

```

```

Align seg 1/1 to: R05159 from: 1 to: 368

1005 AACACAGTGGTGGCGGACGAACACCTGCTGCAGCTGGCGCAACTTGGTTTA 1054
      |||:::||||:::|||||:::|||||:::|||||
86 LysGluLleSerProAspIrrLeuLeuAspLeuGlnAsnAspI 102
1055 CTCA.....AGAAAGAT..... 1067
      ||| ||| ||| |||
102 eserGluLeuArgLysAspPheLysGlyLeuGlnHisLeuTyrAla 119
1068 ....CTGCAGAACCGGAGATATCATGTGATGTACGAAACGATATTCACA 1112
      ||| |||:::|||||:::|||||:::|||||
119 euValLeuValAsnAspLysIleSerLysIleHisGluLysAlaPheSer 135
1113 CCCTTAAATATACGATATATCAATATCAAGCTTGGCATGTGATTAAGCTCAC 1162
      |||:::|||||:::|||||:::|||||:::|||||
136 ProLeuArgAsnVal.....GlnLysLeuTyrIleSe 146
1163 CGTGCACCTTGTCAACATATATGCAGAACCTGTATACTTGGCTGCAGTACA 1212
      :
146 rLys..... 147

```

```

1213 AAGTGAATTCGATGCATGCACATTTGGCTTCACATGTTTATGGGTTAA 1262
148 .....AsnHisLeuValGlu..... 152
1263 GCATTGATTCATGCTGCTGCCGTCATTATTACCATTTGCTTAATTCGCT 1312
153 .....IleProProAsnLeuProSerSerLeu.....Val 162
1313 CCAATTCAGAGGCGATAAATAATGTTAGA.....ATCA 1347
112 ..... 112
162 IASpAlaTrgIleHisAspAsnArgIleArgLysValProLysGlyValP 179
1348 TCGAAGCACA..... 1358
179 heSerGlyLeuArgAsnMetAsnCysIleGluMetGlyGlyAsnProLeu 195
1359 .....ATPAAACTGCCATA 1372
196 GluAsnSerGlyPheGluProGlyAlaPheAspGlyLeuLysLeuAsnTly 212
1373 TATATTCGTTCTTCTTATCATCCTTCTAATACATAATTTAGTACACA 1422
112 ..... 112
212 rLeuArgIleSerGluAlaLys.....LeuThrG 222
1423 ATATATAGTATAGTAGAATAATTAATCTGTAATACATAAGTATATTTTC 1472
222 Lylle.....ProLysAspLeuProGluThrLeuAsnGluLeuHis 235
1473 ATCAAAACTTCTTATGCGCTTTTATAGCTTGTGAAGCTTATTCATTA 1522
236 LeuAspHisAsnLysIleGlnAlaIleGluLeuGluAspLeuLeuArgTly 252
1523 TTCAGTAATCTTTTATATGCAATATATGTAATGTTTCATCATAGGCC 1572
252 rSerLysLeuTyrArgLeu.....GlyLeuGlyH 262
1573 ATGAACTATTCGTTGTTTATATGCA...TCATATCACTTGTCTCT... 1616
262 ISAsnGlnIleArgMetIleGluAsnGlySerLeuSerPheLeuProThr 278
1616 ..... 1616
279 LeuArgGluLeuHisLeuAspAsnLysLeuAlaArgValProSerG1 295
1617 .....ATTATATCTTAACAGTTTATCATTTGTGATATA 1650
295 YLeuProAspLeuLysLeuLeuGlnValValTyrLeuHisSerAsnAsnI 312
1651 TCACAATTTATACCTTG.....TATGCCCCAATTTTATGGGCATC 1691
312 IeThrLysValGlyValAsnAspPheCysProMetGlyPheGlyVal 327
seq_name: A_Geneseq_36:R88469
seq_documentation_block:
ID R88469 standard; Protein: 1464 AA.
AC R88469;
DT 14-AUG-1996 (first entry)
DE Feline infectious peritonitis 1 virus spike protein.
KW Feline infectious peritonitis 1 virus; FIPV-I; spike protein;
OS vaccine; prevention; treatment.
OS Feline infectious peritonitis 1 virus.
PN J07327683-A.
PD 19-DEC-1995.
PE 10-JUN-1994; 129300.
PR 10-JUN-1994; JP-129300.
PA (KITA ) KITASATO KENKYUSHO SH.
DR WPI; 96-072341/08.
DR N-PSDB; T10166.
PT DNA encoding feline infectious peritonitis 1 virus spike protein -
used in a vaccine for prevention and treatment of FIPV-I infection
PS Claim 1; Page 14-17; 23pp; Japanese.

```

CC This sequence represents the feline infectious peritonitis 1 virus
 CC (FIPV-I) spike protein. The FIPV-I spike protein may be used in the
 CC production of a vaccine for the prevention and treatment of FIPV-I
 CC infection. The spike protein may be produced by transforming a host
 CC cell with the spike protein DNA and expressing the sequence such
 CC that the spike protein can be isolated.
 SQ Sequence 1464 AA;

alignment_scores:
 Quality: 90.50 Length: 382
 Ratio: 0.508 Gaps: 16
 Percent similarity: 46.597 Percent identity: 19.895

alignment_block:
 US-09-323-427-1 x R88469 ..

Align seg 1/1 to: R88469 from: 1 to: 1464

```

299 CGTAATGCAATTCGAGACATGTTATGTGAAA.....GGTCTTTA 339
:::||||: |||:||||:|
739 LysAsnSerThrThrGlyGluIlePheThrValProCysAspLeuH 755
:::||||: |||:||||:|
340 rGATCAGAAGGTCGCCGTAATGATGAAGTGACGTCAAGTTGCCGGA 389
|||: |||:||||:|
755 rAlaGlnAlaIleValIleAsnAspGluIleValGlyAlaIleThrAlaV 772
::: |||:||||:|
390 TTTCACTTCCA.....TTTGATTCATGCAAT.....GTGCCGT 424
|||: |||:||||:|
772 AlAsnGlnThrAspLeuPheGluPheValAsnAsnThrGlnIleArgH 788
|||: |||:||||:|
425 ACACGATCTGTAATCAACGCTGATTTTGTACACAACTGTGTGATC 474
:::||||: |||:||||:|
789 SerArgSerSerThrProAsn.....PheValThrSerTyrThrMetPr 803
:::||||: |||:||||:|
475 TTCGTTTCATCCATTATTTGTTACCAAGTTGATGCGATTCAGATAC 524
:::||||: |||:||||:|
803 oGlnPheTyr.....TyrIleThrLysTrpAsnAsnSprThrSerSerA 818
:::||||: |||:||||:|
525 AATGCTTTTACATGAGAGCTGATTAACAGTGAAGTGCACAGATGAGTA 574
:::||||: |||:||||:|
818 snCys.....ThrSerAlaIleThrTyr 825
:::||||: |||:||||:|
575 TCTGAATTCACAACCTGCTTTTCACAACCTCAATTTGCCGATCCAGTATG 624
|||:||||: |||:||||:|
825 SerSerPheAlaIleCysAsnThrGlyGluIleLysTyrValAsnValHn 842
:::||||: |||:||||:|
625 CCGTATGCAATTTTGATGCT..... 646
::: |||:||||:|
842 rHisValGluIleValAspAspSerIleGlyValIleLysProValSerT 859
::: |||:||||:|
647 ..GGACCAACCGCTACACCGTCAATTTGCTATCATTTGTCGACAGAT 694
|||: |||:||||:|
859 rGlyAsnIleSerIleProLysAsnPheThrValAlaValGlnAlaGlu 875
::: |||:||||:|
695 TATCAATAATGACATGCGATTCGTGAACCGTTGATCTTCTGCGCGGT 744
|||: |||:||||:|
876 TyrIleGlnIleGlnValLysProValValAsp.....CysAlaLeh 890
::: |||:||||:|
745 TGTCAATTCCTGCTTTGCGATGATGTAACGGTGATCTGGAATTC 794
::: |||:||||:|
890 r.....TyrValCysAsnGlyAsnThrHisCysLeuLysLeuL 903
::: |||:||||:|
795 Tr.....AATGCT.....GAT 805
|||: |||:||||:|
903 eutHrGlnTyrThrSerAlaCysGlnThrIleGluAsnAlaLeuAsnLeu 919
::: |||:||||:|
806 CGATGCTCTTGATTAATATTGCTAAATATTTGATATCAACAGAGA 855
|||: |||:||||:|
920 GlyAlaArgLeuGlnSerLeuMetLeuAsnAspMet..... 931
::: |||:||||:|
856 TTTAATGCTGCCCAAGAGCTCACGTATACAAATATGCGGATGCATCAC 905
::: |||:||||:|

```

```

931 ..... 931
906 AGCTTTCTATCAATGCAGATCATATTACATTAAGACCAATAGC 955
932 ..... 939
956 GAATGTGTTGACACAGATGTTTCAGACCAAGAGATTGCGAGCTTTAA 1005
940 GtuleuAlaThrValGluArgPheAsnAlaThrAlaLeuGlyGly 956
1006 AACAGGTGTCGCCGACGACCAAAACCTGCTGACGTGCCCAATTCGTTTAC 1055
956 stleuGlyGlyLeuTyrPheAspGlyLeuSerLeuLeuProLys1 973
1056 TCAGAAAAGATGTCGAGAACCGAGATATATCATGTAT.....GTACGA 1099
973 leuGlySerArgSerAlaValGluAspLeuPheAsnLysValValThr 999
1100 ACTGATATCAACACCCCTTGAAATTCAGATGATATCAACCTTGGCAGT 1149
990 SerGlyLeuGlyThrValAspAspAspTyrLysLysCysSerSerGlyThr 1006
1150 TGATTTACGTCACCGTCACCTTCTGCACATATGACAACTGTATATC 1199
1006 rAspValAlaAspLeuValCysAlaGlnTyrTyrAsnGlyIleMetVal1 1023
1200 TTGCTGCAGTACAAATGGAATCTGCATGTCACCATTTGGCTTCTCAAG 1249
1023 euProGlyValValAspGlyAsnLysMetSerMetTyrThrAlaSerLeu 1039
1250 TTTATGGGTTTAAGCATT...GCATTGATTGCTGCCGTCATTATTACCAT 1296
1040 IleGlyGlyMetAlaLeuGlySerIleThrSerAlaValAlaVal1.ProP 1056
1297 TT.....CGTTAAATTTGCTCCAAATCAGA 1322
1056 heAlaMetGlnValGlnAlaArgLeuAsnTyrValAlaLeuGln 1070

seq_name: A_Geneseq_36:W88310
seq_documentation_block:
ID W88310 standard; Protein: 374 AA.
AC W88310.
DT 26-APR-1999 (first entry)
DE E. coli colitoxin or glucose transferase.
KW O antigen; O11 antigen; wbdm gene; colitoxin transferase;
KW glucose transferase; diarrhoea; haemorrhagic colitis; diagnosis.
OS Escherichia coli.
PN M09850531-A1.
PD 12-NOV-1998.
PE 01-MAY-1998; AU0315.
PR 22-JUL-1997; AU-008162.
PR 01-MAY-1997; AU-006545.
PA (UNSY ) UNIV SYDNEY.
PI Reeves PR; Wang L;
DR WPI: 99-059669/05.
DR N-PSDB: X06748.
DR Nucleic acid molecules specific for bacterial polysaccharide
PT antigens - useful for detecting specific strains in, e.g. food,
PT faeces or patient samples
PS Disclosure: Fig 7; 165pp; English.
CC This is the amino acid sequence of the protein encoded by the wbdm
CC gene of a gene cluster (see X06748) involved in the biosynthesis of
CC the Escherichia coli O11 antigen. The protein shows high
CC homology with T8E of Yersinia enterocolitica, and is predicted to
CC be a colitoxin or glucose transferase. The use of nucleic acid
CC molecules derived from particular assembly and transport genes,
CC particularly wbd (transferase), wxz (flippase) and wzy (polymerase)
CC genes, within O antigen gene clusters improves the specificity of
CC methods for the detection and identification of O antigens, e.g. in
CC testing food- or faecal-derived samples, or samples from patients.
CC The O antigen is a major virulence factor of enteropathogenic E.
CC coli strains that cause diarrhoea and haemorrhagic colitis.

```

```

SQ Sequence 374 AA;
alignment_scores:
  Quality: 89.50      Length: 363
  Ratio: 0.533
  Percent Similarity: 46.281      Percent Identity: 21.763
alignment_block:
US-09-323-427-1 x W88310
Align seg 1/1 to: W88310 from: 1 to: 374
228 TTGACAATGCTGTGCAAGGTGAGCCGAAATTCATGATGACCACTTCA 277
38 leuThrGlyIleSerGluValLysProThrGlnAsnIleAsn..... 51
278 ATACATCAATTTTATATACAGTAAATGCAATTCGAAAGAC.....A 318
52 .....IleHisTyrValAsnMetLysPlysAsnPheArgS 63
319 TGTATTGTAAGAAGCTTTATGATCAAGAGGTGCGCTTAATGATGAG 368
63 eTrPheArgAlaLeuPheGlnValLysLysIleIleValAlaLeuLys 79
369 GTGACGTCAGCTTGGCGGAATTCACCTTCATTTGATTCATGCAATGT 418
80 ProAspIle.....IleHisSerHisMetPheHisAlaAsn.. 91
419 GCGCGTCACGATCTGTGATCCAGCTGATATTTTGTACAAACAACGT 468
92 .....I 92
469 TGTGATTTGTTTATCCATTAATTTGTTTACCAAGTTG...ATCGTGCAT 515
92 lePheSerArgPheIleArgMetIleIleProAlaValProIleuLecys 108
516 ATCGAGTCAATGCTTTTACATGGAAGCTGATTAACACTGTAGTCACAG 565
109 ThrAlaHisAsnLysAsnGlnClyLysAsnAlaArgMetPheCysTyrTr 125
566 ATTGACGATATCGAATCACAACGCTTTTCAACGCAATATGTCGCCGAT 615
125 glu.....SerAspPheLeuAlaSerIleThrThrAsnValS 138
616 GCGAGTATCCGTTATGAAATTTGGATGTGACCAACGCGTCAACGAG 665
138 erLysGluAlaValGlnGluPhe.....IleAlaArgLysAlaThrPro 152
666 TTCATTTGCTATCATGTGTCAGCCAGTTTATCAT..... 700
153 LysAsnLysIleValGlnIleProAsnPheIleAsnThrAsnLysPheAs 169
701 .....AAATGACATGCGATTCGTGAACCGTTGATA 731
169 pheAspIleAsnValArgLysLysThrArgAspAlaPheAsnLeuLysA 186
732 CTTTTCGCCGGTGTGTCATCTGCTTGTGATGATGATGAACGATGAT 781
186 spSerThrAlaValLeuLeuAlaVal.....GlyArg 196
782 ACTGTGGAA.....ATTCTAATGCTGATGATGTGC 813
197 LeuValGluAlaLysAspTyrProAsnLeuAsn.....Al 209
814 TCTGTATAATATTGCTAATAATTGCAATATCCACAGAT..... 856
209 alleAsnHisLeuIleLeuSerLysThrSerAsnCysAsnAspPheIle 226
857 ..TTATAGCTGCGCAAGACTCACTGATACCAATATGCGGATGATCA 904
226 euLeuLeuIleAlaGlyAspGlyAlaLeuArgAsnLysLeuLeuAsp..... 240

```

```

905 CAGCTTTCTATCAATGCCAGATCAGTATACCATTA..... 943
241 .....LeuValAlCysGlnLeuAsnLeuValAspLysValPhePhe 254
944 ..GAACCAATAGCGAATGTCTTCGACCACAAATGT..... 976
254 uGlyGlnArgSerAspLysGlnLeuMetCysAlaAlaAspLeuPheV 271
977 .....TCGAACCAACAGGATTCGGA...GCTGTAAACAGAGTGT 1015
271 allLeuSerSerGluTrpGlnGlyPheGlyLeuValAlaGlnLamet 287
1016 GCCCGACGAAACCTGCTGCAGCTGCGCAACTGCTTACTCACAAG 1065
288 AlacysGluArgProValValAlaThrAspSerGlyGlyValGluVa 304
1066 ATTCGCACACCGGAGATATCATGTAGTATGACATGATATC..... 1108
304 lValGlyProHisAsnAspValLleProValSerAsnHisLleLeuVa 321
1109 .....AACACCTTGAAATTAAGCATATATCAACGCTTTG 1144
321 lalGlyLleAlaGlnThrLeuLysLle...AspAspAsnAlaArgLys 336
1145 CCAGTTCATTATACGTACCGTGCACCTTCTGCAACATAAT 1183
337 lIleGlyMetLysAsnArgGlnTrpLleValSerAsn 349

seq_name: A_Geneseq_36:R59926

seq_documentation_block:
ID R59926 standard; protein; 3079 AA.
AC R59926;
DT 22-FEB-1995 (first entry)
DE GAP protein Iira2.
KM Ras: GTPase activating protein; GAP: GAP related domain; GRD;
OS Saccharomyces cerevisiae.
PN MO9416069-A.
PD 21-JUL-1994.
PF 12-JAN-1994; 000198.
PR 15-JAN-1993; US-004824.
PA (SCHE ) SCHERING CORP.
PI Kaziro Y, Nakafuku M.
DR MPI; 94-249216/30.
PT Blocking Ras-induced effects on a cell - by introducing a GTPase
PT activating protein to the cell, used esp. in treatment of cancers
PS Disclosure; Page 63-72; 87pp; English.
CC Human neurofibromatosis type 1 (NF1)-GAP related domain (GRD)
CC mutant clones NF201 (given in R59221) and NF204 (R59922) show
CC v-Ras-induced transformation in mammalian cells. The mutation
CC sites of these proteins were located in one of the most conserved
CC regions of GRD. These sites were compared with those of other
CC GPD family proteins, yeast Iira2 (R59926) and Irai (R59923).
CC human GAP (R59924) and Schizosaccharomyces pombe Gap1 (R59925).
SQ Sequence 3079 AA.

```

alignment_scores:

```

Quality: 89.00 Length: 752
Ratio: 0.264 Gaps: 42
Percent Similarity: 44.814 Percent Identity: 18.617

```

alignment_block:

```
US-09-323-427-1 x R59926 ..
```

```
Align seg 1/1 to: R59926 from: 1 to: 3079
```

```

38 ACTGCAGCAACAACAACAAC.....AAACAACAACAACAACAAC 81
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
2300 ThirSerSerProAsnSerAsnAsnLysValLysGlnGlnLysGluArgTh 2316

```

```

82 AGCAATATATACCCCATCAATGAGGAGGAAGACAGCAAGCAATCTAG 131
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
2316 lIleLeuLeuCys.....HisLleLeuLeuV 2325
132 TTTTCTATAAATCGAATTTACTAATCTTCTGAATG.....ATG 172
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
2325 alSerLeuLleGlyLeuPheAspGlnSerLysLysMetLysAsnSerSer 2341
173 ATTCGCTTATATGCTTCTGCTACACTTATTCGATTCGCTTATTCG.. 220
2342 TyrAsnLleLleAlaAlaThrGlnLalaserPheGlyLeuAsnPhelYse 2358
221 .....ATTCCGGTTCACAAT..... 235
2358 rHisPheHisArgSerProGlnValTrpValProGlnAspThrThr 2375
236 .....GCTTCGAAGTGAAGCCAGAAATTCGATTCGAGCAATTCATA 280
2375 heLeuGlyValLleGlyLysSerLeuAlaGlnSerAsnProGlnLeuThr 2391
281 ACAATCAATTTTAATACAGGTATGCATTCG...AAGGACATGTTATGT 327
2392 AlATrMetPhe.lIeTrpValLeuGlnAlaLeuLysAsnAsnValLleP 2408
328 GAAGGCTCTTATATGATCAAGAAGTTGCC..... 356
2408 rOniValTrpLleProHisThrLleCysGlyLeuSerTrpTrpLlePro 2424
357 .....GTAAATGATGAAGTGAAGTCAGTCAGTCCGGAAT 391
2425 AsnLeuTrpGlnHisValTrpLeuAlaAspAspGlnGlnLysProGlnAs 2441
392 TCACCTTCATTTGATTCATGCAATGTCGCC.....GTACAGCATCTCT 435
2441 nLleSerHisLlePheArgGlnLeuLleArgLeuSerValAlaGlnLuthA 2458
436 GAATCCAGCTGATATTTTGTACAAACACTGTGTCATTT..... 476
2458 sPrheLysAlaValTrpMetGlnTrpValTrpLleLeuLeuLeuAspAsp 2474
477 ...GGTTCATCATATTTGTTACCAAGTGCATGTCGCAT...ATGCA 520
2475 GlnArgLeuThrAspLleLleValAspGlnValLleAsnHisAlaLeuL 2491
521 GTCAATGCTTTTACA.....TGGAG..... 542
2491 uArgAspSerGlnAsnArgAspTrpLysThrLleSerLeuLeuThv 2508
543 .....CTGATAAACAAGTTAGTGCAC 563
2508 allLeuProThrThrGlnValAlaAsnAsnLleLleGlnLysLleLeuAla 2524
564 AGATGAGGATATCTGAATACAACTGCTTTCAACATCAATTTGTCGG 613
2525 LysLleArg.....SerPheLeuProSerLeuLysLeuGln.. 2536
614 ATGCCAGTATGCCCTTATGCAATTTGTGAGTGAGCAACCGGTCAAC 663
2537 .....AlaMetThrGlnSerTrp.....Serg 2544
664 AGTTCATATTTGCTATCATTTGTCAGCCAGTTTATCATATAATGAGACATCG 713
2544 luleuThrLleLeuValLysLleSerLleHisValPhePheGlnThrSer 2560
714 ATTCGAAACCGTGCATCTTCGCGCGCTTCGATTCCTTCGCTTGTGC 763
2561 leuLeuValGlnMetLysLeuProGlnLleLeuPheLle.....ValSe 2575
764 GATGATGTGAACGTGATACCTGTGAATTTCAATTCGATGATGATGTC 813
2575 rleuLeuLleAspValGlnProArgGlnLeuArgSerSerLeuHisGlnL 2592
814 TCTTGATAATAATTTGCTAATAATTTGCAATATATCAACAGATTTAATGC 863

```

```

||||:||||: 2597
euleumetalsnValCys..... 2592
CTGGCAGAGACCTCAGCATACAAATATGGCGATGATCAGAGCTTTTC 864
.....HisSerLeuAl 2601
914 TATCATGCGACATGATATTACATTAAAG.....AMCC 948
||||:||||: 2601
alleasn.....SerAlaleuProGlnaspHisarGasnasnLeuAspG 2616
949 AAATGCGCAATGTGTCGACCAATGTTCCAGAAC..... 983
||||:||||: 2616
IuileSeraspIlePhealahlSglnLysvalLysPheMetPheGlyPhe 2632
984 CACAAGATTCGGAGCTGTTAAACAGGCGGCCGAC..... 1022
||||:||||: 2633
SerGlnaspLysGlyArgIleLeuGlnIlePheSerAlaSerSerPheAl 2649
1023 .....CAAACTGCTGCAGCTGC 1041
|||:||||: 2649
aserLysPheasnIleLeuaspPhePheIleasnIleLeuLeu 2666
1042 GCAACTTC.....GTTACTCAAGAAAAGATCTCAGAACCGGAGAAATA 1085
||||:||||: 2666
eGluTyrSerSerThrTyrGlnAlaasnValTrpLysThrArgTyrLys 2682
1086 TCATGATGATGACACTGATATACACA.....CCCTTGAA 1120
||||:||||: 2683
LysTyrValLeuGlnSerValPheThrSerAsnSerPheLeuSerAlaAr 2699
1121 ATTACGATGATTAATCAAGCTTTGCCAGTGAATTACGTCACCGGCACCT 1170
|||:||||: 2699
gSerIleMetIleValIleGlyIleMetGlyLysSerTyrIleThrGlnGlyL 2716
1171 TCTGCAACATATG.....GACAACGCTGATACTGTCGAGTAACAA 1214
||||:||||: 2716
eCysLysAlaMetLeuIleGlnThrMetLysValIleAlaGlnProLys 2732
1215 ATGGAATCT.....GCATGTCACCAT..... 1235
||||:||||: 2733
IleThrAspGlnHisLeuPheLeuAlaIleSerHisIlePheThrTyrSe 2749
1236 .....TTGGCTTCTCATATGTT 1251
2749 rLysIleValGlnGlyLeuaspProAsnLeuaspLeuMetLysHisLeuP 2766
1252 TATGGGTTTAAGCATTG.....CATTGATTGCTGCCGCTCA 1286
||||:||||: 2766
hetTrpPheSerThrLeuPheLeuGlnSerArgHis.....ProIle 2779
1287 TTAATTACCATTTCTTTAAATTTCTGCCAA.....ATCAGAGG..... 1325
||||:||||: 2780
IlePheGlnGlyAlaLeuLeuPheValSerAsnGlyIleArgArgLeuTyr 2796
1326 .....CATAAATAATGTTAGATTCATCGAAGCAATA 1359
||||:||||: 2796
rMetAlaGlnPheGlnAsnGlnSerGlnThrSerLeuIleSerThrLeu 2813
1360 TA.....AACTGCCATATATATCGTTCTTCTATATACCTTCT 1400
||||:||||: 2813
eULysGlyArgLysPhealahlStnrrPheLeuSerLysIleGlnAsn... 2828
1401 AATACTAATATTACATAACAATATATAGTATGAGCAAAATAATTACTG 1450
||||:||||: 2829
LeuSerGlyIleValITrpaSngLusAspAsnPhen 2840
1451 TAATACA.....ATAAGCATATATTTCARCA 1476
||||:||||: 2840
rHisIleLeuIlePheIleIleAsnLysGlyLeuSerAsnPropheIle 2857
1477 AAACCTTCTTCTATCGCTTTTATA..... 1499
||||:||||:

```

```

2857 ySSerThrAlaPheaspPheLeuLysMetMetPheArgAsnSerTyrPhe 2873
1500 .....GCTTGAAAGCTTATTCATTAT...TCAGTAATCTTTTA 1537
||||:||||: 2874
GlnHisGlnIleAsnGlnLysSeraspHisThrLysCysTyrMetPhe 2890
1538 TATGATACTATTGTAATGTTTCATCA.....TTAGGCCATG 1575
||||:||||: 2890
uLeuTyrPheValLeuAsnGlyAsnGlnPheGlnGlnLeuLeuGlnLysAspV 2907
1576 AATAGTTTCGT.....TTGTTATATATCATCATTTCAACTGCTCT 1616
||||:||||: 2907
aLaspPheGlnGlyGlnMetValasnIleGlnAsnLysasnThrIlePro 2923
1617 ATTTATTTCTAAGCTTTATCATTTGTGATTAAT...ATCAGCAATATTATAC 1663
||||:||||: 2924
LysIleLeuLeuGlnTrpLeuSerSeraspAsnGlnLusAlaAsnIleTh 2940
1664 CTGTAT.....TGCCCAATT..... 1679
||||:||||: 2940
rLeuTyrGlnGlyAlaIleLeuPheLysCysSerValThrAspGlnPro 2957
1680 .....TTTATGGGCACTTTCCTATTCCTGTAACAAATTCACCTT 1718
||||:||||: 2957
eArgPheArgPheAlaLeuIleIleArgHisLeuLeuThrLysLysPro 2973
1719 ATTTCATTAATTGCAATTAAAGATTTTCATTGTCGAAAAAATA 1768
||||:||||: 2974
IleCys.....AlaLeuArgPheTyrSerValIleArgAsnGlnIleAla 2988
1769 AAAA 1772
||||: 2988
gLys 2989

seq_name: A_Geneseq_36:R42452
seq_documentation_block:
ID_R42452 standard; Protein; 1634 AA.
AC_R42452;
DE 27-MAY-1994 (first entry)
DE Enzyme involved in eicosapentaenoic acid (EPA) synthesis.
KW EPA; eicosapentaenoic acid synthetase; drug; anticoagulant;
KW hypolipemic; hypoglycemic; antihypertensive; anticancer; pesticide;
KW foodstuff; additive.
OS Shewanella putrefaciens.
PN MO9323545-A.
PD 25-NOV-1993.
PE 14-MAY-1993: J00641.
PR 15-MAY-1992: JP-147945.
PA (SAGA ) SAGAMI CHEM RES CENTRE.
PI Kato S, Kondo K, Yamada A, Yazawa K;
DR Kato S, Kondo K, Yamada A, Yazawa K;
DR N-PSDB: 051128.
PT Gene coding for eicosapenta:enoic acid synthetase - is isolated
PT from Pseudomonas, Alteromonas or Shewanella and used for
PT recombinant prodn. of eicosapenta:enoic acid
PS Claim 6; Page 52-63; 106pp; Japanese.
CC EPA is useful as a drug, having antitumor, hypolipemic,
CC hypoglycemic, antihypertensive and anticancer activity. It is also
CC a pesticide and is useful as a nutritional foodstuff and animal feed
CC additive.
SQ Sequence 1634 AA.

alignment_scores:
Quality: 88.00 Length: 500
Ratio: 0.376 Gaps: 23
Percent Similarity: 46.800 Percent Identity: 18.200

alignment_block:
US-09-323-427-1 x R42452 ..
Align seg 1/1 to: R42452 from: 1 to: 1634

```

Align seg 1/1 to: US-08-415-751-6 from: 1 to: 362

```

790 TTCACAGTATCACCGTTACATCATCGAACAAGCAGAAATGAC..... 746
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
22 PheHISserLeuSerHISleuIleuAspLysGlnPheThrSerPheTr 38
745 .....AACCGCGAAGATATCAGCGTTTCAGATATCGATGTC 706
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
38 PLeuIleuArgAsnGlnThrAsp..... 45
705 CATTTATGATTAACGGTGGACCATGATGCAATGTAAGTGGTGGACC 656
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
46 ....TrrPLeuIleuTrpValThrSerGluCys...IleAsnTrp**thr 59
655 GGTGTGTCACCATCCAA.....AATTTCATACGGCATACTGGCAGTCG 612
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 ValValSerSerValSerThrValLysPheLeuAlaIleArgTyrAspPh 76
611 GGACAATTTGAGTTGAAAGACGATTGTGATTTTCAGATACCTCAATCTGT 562
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
76 rAspLysProAspIle..... 81
561 GCACTAAGTGTATTATCAGCTTCATGTAAGAAGCATGTGATGATATGC 512
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
82 .....Val**LeuValIleGly...Cys 88
511 ACG.....ATCACTTGTGTATC.....AATAATGATGTAAGCAAGAA 474
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89 ThrLysLeuProAspPheGlySerAsnGlyArgProThrIleGlySerI 105
473 TGACACAGCTGTGTGTACAAAAATACACGCGATTCAGAGATGCTGTA 424
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
105 eGlyLysProValLeuLeu**Val**ThrTrpGlu**IleAlaCysI 122
423 CGCGCAATTCATGATGAATCANA...TGAAGTGAATTCGCGCAACTTG 377
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122 LeSerTyrLeuSerCysLeuArgThrTrp..... 131
376 ACGTCCACCTTCATCATT.....ACGGCAACCTTCGTGATCATATAA 336
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
132 ...AsnTrpPheIleIleArgVal*******ThrThrValIleVal.. 146
335 GACCTTCACATAAACATGTCCTTCGATGCAATTCAGTGTATTAATG 286
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
147 .....IleAsnHISer.....CysValLeuIleG 155
285 ATTGTTATTGAAGTGGTCCACATCATTCATTTCTGGCTCACCTTCACAC 236
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
155 InHISLeuPheAsnTrpAspTrpCysAsnPhe..... 165
235 ATTGTCAACCGGAATGAAATAGACAATGCAATAAGTGTAGTACGAAGA 186
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
166 .....AsnValThrAspIleArg**TrpCysArgCysCys..... 178
185 CAATTAAGACGAATCATTCATTTCAGAGAAGATTAGTAATTCGATTTTGA 136
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179 .....HisPhe..... 180
135 AAAACTAGATTGCTCTCTCTCTCTCCACATTGATGGGATATTAT 86
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 .....ValSerLeuGlySerSerSerCysArgTrp...HisCys 192
85 TGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 47
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
193 CysCysCysCysCysCysCysPheCysCysCysCysCysAsnTyrTrpLe 209
46 ....TGCTGCAGTCGCTGT 32
209 u**CysCysCysCysCys 215

```

seq_name: /cgn2_6/plodata/1/aa/5a_COMB.pep:US-08-415-751-35
seq_documentation_block:

```

? Sequence 35, Application US/08415751
? Patent No. 5643772
? GENERAL INFORMATION:
? APPLICANT: PETERSEN, CAROLYN
? APPLICANT: LEECH, JAMES
? APPLICANT: NELSON, RICHARD, C.
? APPLICANT: GUT, JIRI
? TITLE OF INVENTION: POLYPEPTIDES BINDING ANTI-
? TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA
? TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID
? TITLE OF INVENTION: VECTOR AND TRANSFORMED HOST AND
? TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND
? TITLE OF INVENTION: DIAGNOSIS AND KIT
? NUMBER OF SEQUENCES: 50
? CORRESPONDENCE ADDRESSES:
? ADDRESSEE: PHILLIPS, MOORE, LEMPIO & FINLEY
? STREET: 385 Sherman Avenue, Suite 6
? CITY: Palo Alto
? STATE: California
? COUNTRY: United States of America
? ZIP: 94306-1840
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Kb storage
? COMPUTER: PC
? OPERATING SYSTEM: DOS
? SOFTWARE: Wordperfect 5.1
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/415,751
? FILING DATE: 03-Apr-1995
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/071,880
? FILING DATE: June 1, 1993
? APPLICATION NUMBER: 07/891,301
? FILING DATE: May 29, 1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Hana Dolezalova
? REGISTRATION NUMBER: 30,518
? REFERENCE/DOCKET NUMBER: 480.19-2 (HHD)
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (415) 324-1677
? TELEFAX: (415) 324-1678
? INFORMATION FOR SEQ ID NO: 35:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 362 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: peptide
? ORIGINAL SOURCE:
? ORGANISM: Cryptosporidium parvum
? FEATURE:
? NAME/KEY: Positions coded by nonsense codons are
? NAME/KEY: identified as Xaa.
? US-08-415-751-35

```

alignment_scores:

Quality:	109.00	Length:	31
Ratio:	6.056	Gaps:	0
Percent Similarity:	58.065	Percent Identity:	51.613

alignment_block:

US-09-323-427-1/rev x US-08-415-751-35 ..

Align seg 1/1 to: US-08-415-751-35 from: 1 to: 362

```

124 TGCTTCTGTCTCTCTCTCCACATTGATGGGTTATTGCTGTGTG 75
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
135 CysCysCysCysLeuAsnTrpThrLeuTrpTrpTyrGlyCysCysCys 151
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
74 TTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 32
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```


; MOLECULE TYPE: protein

692 [vAspD]aH[sA]aAsnA]

	Quality:	100.00	Length:	391
	Ratio:	0.541	Gaps:	18
	Percent Similarity:	47.315	Percent Identity:	19.437
<p>mmnt_block:</p> <p>09-323-427-1 x US-07-707-367-2 ..</p> <p>ign seg 1/1 to: US-07-707-367-2 from: 1 to: 935</p>				
263	TTGTGGACCACTTCATTAACATCATTTTAAATACAGTAATGCATTGCA	312	TTGTTGACCACTTCATTAACATCATTTTAAATACAGTAATGCATTGCA	312
526	CysdlnProthrsprAlaIaValasProllethnglnaPhepheva	542	ProthrsprAlaIaValasProllethnglnaPhepheva	542
313	AGCAGCATGTTTATGTGAAGAGCTTTATGATCAAGAGCTTCCGCGATG	362	AGCAGCATGTTTATGTGAAGAGCTTTATGATCAAGAGCTTCCGCGATG	362
542	lAlaasprGlyTyr.....	546	lAlaasprGlyTyr.....	546
363	ATGAGAGTGGAGCTCAAGTTGCCGGAAATTCACATTCATTGATTATGC	412	ATGAGAGTGGAGCTCAAGTTGCCGGAAATTCACATTCATTGATTATGC	412
547	547	547
413	AATGTTCCGCGTACAGATCTCGAATCCAGCGTATTTTGTGA.....	457	AATGTTCCGCGTACAGATCTCGAATCCAGCGTATTTTGTGA.....	457
548	Asn...SerArgIleMetcIlnPheSerProasngIlyMetPheIleMetcI	566	Asn...SerArgIleMetcIlnPheSerProasngIlyMetPheIleMetcI	566
457	457	457
563	ntrpGlygluIuThrsSerAsnValaProargProgluInPheargI	580	ntrpGlygluIuThrsSerAsnValaProargProgluInPheargI	580
458ACAACAAGCTGTGTCATTTGGTTTCATCCATTGTTTATACC	499ACAACAAGCTGTGTCATTTGGTTTCATCCATTGTTTATACC	499
580	IeproHisSerLeuThrMetcAlaProasprGlnGlyInLeuCysValaIa	596	IeproHisSerLeuThrMetcAlaProasprGlnGlyInLeuCysValaIa	596
500	AAAGTTGATCGTGCATATGAGATACAAATGCTTTTACATGACAGCTGTAA	549	AAAGTTGATCGTGCATATGAGATACAAATGCTTTTACATGACAGCTGTAA	549
597	AspargluAsnGly...ArgIleGlnCysPheHisAlaIuThrGlyAs	612	AspargluAsnGly...ArgIleGlnCysPheHisAlaIuThrGlyAs	612
550	AACAGCTAGTGGCAGATGTAGAGTATCTGAAATC...ACAAGCTCTTTC	596	AACAGCTAGTGGCAGATGTAGAGTATCTGAAATC...ACAAGCTCTTTC	596
612	nPheValuLys...GlnIleLysHisGlnIlnPheIuArgIuAlaPhea	628	nPheValuLys...GlnIleLysHisGlnIlnPheIuArgIuAlaPhea	628
597	AAACTCAAAATGTGCCGATAGCCAGATAGCCGTTATGAATTTTGATAGGT	648	AAACTCAAAATGTGCCGATAGCCAGATAGCCGTTATGAATTTTGATAGGT	648
628	IaValasprGlyAlaPro.....Gly	634	IaValasprGlyAlaPro.....Gly	634
647	GGACCAACCGCGTACACCAAGTCAATTTGCTATGCTATGCTACGCCAGTTA	696	GGACCAACCGCGTACACCAAGTCAATTTGCTATGCTATGCTACGCCAGTTA	696
635	Gly.....ValLeuArgIaValasInGlnLysProTyrTy	646	Gly.....ValLeuArgIaValasInGlnLysProTyrTy	646
697	TCATTAATGAGCATGCGGATTCGAAACGTTGATACTTTCGCCGGTTG	746	TCATTAATGAGCATGCGGATTCGAAACGTTGATACTTTCGCCGGTTG	746
646	r.....GlyIleSerIaIaPro	652	r.....GlyIleSerIaIaPro	652
747	TTCATATCTGCTGTTTGTGATGATGCTAAAGGTATGCTGCGAAATTTCTA	796	TTCATATCTGCTGTTTGTGATGATGCTAAAGGTATGCTGCGAAATTTCTA	796
652	alGlnGlyPheMetcLeuAsnPheSerAsnGlyAsp.....	663	alGlnGlyPheMetcLeuAsnPheSerAsnGlyAsp.....	663
797	AATGCGATGAGATGCTGCTGTATTAATATTG.....CTAATTAATT	840	AATGCGATGAGATGCTGCTGTATTAATATTG.....CTAATTAATT	840
664IleLeuAspThrPheIleProAlaIaGlyAsnPrh	675IleLeuAspThrPheIleProAlaIaGlyAsnPrh	675
841	GGAATATCCACAGATTTTAATGGCTGGCCAAAGAGCTACAGTATCAAT	890	GGAATATCCACAGATTTTAATGGCTGGCCAAAGAGCTACAGTATCAAT	890
675	eAspMetProHisAsprIleAlaIaIaIaAsprAspIuThrsValuIaValG	692	eAspMetProHisAsprIleAlaIaIaIaAsprAspIuThrsValuIaValG	692
891	ATGCGAGTGCATGACACAGCTTTTCTATCAATGCCAGATCAGT.....	931	ATGCGAGTGCATGACACAGCTTTTCTATCAATGCCAGATCAGT.....	931
992	IuAsprAlaHisAlaAsnAlaValIuTrpLysPheSerProSerIlyValaGlu	708	IuAsprAlaHisAlaAsnAlaValIuTrpLysPheSerProSerIlyValaGlu	708


```

597 AAACCAATGTCGCCGATGCGCATATGCCGTATGAAATTTGGATGCT 646
      ::::: ::::: |||
744 lavalserTyrAlaPro.....Gly 750
647 GGACCAACCGGTCAACCAATTGCTATCATCTTGGTCAGCAGTTTA 696
      ||| ::::: |||
751 Gly.....ValLeuTyrAlaValasnGlyLysProTyrTy 762
697 TCATTAATGACATGCATTCCTGAAACCGTTGATCTTGGCGCGGTG 746
      | ||||| ||::: |||
762 rGly.....AsperThrProValGlnGlyPhe..... 771
747 TCATTCCTGCTTTGTCGATGATGATGACGATGATCTGGAATTTCTA 796
      ::::: ::::: |||||
772 .....MetIleasnPheSerasnGlyAsp..... 779
797 AATGCTGATGATGCTGCTTGTGATAATATTG.....CTAATAATT 840
      ||||| ::::: :::::
780 .....IleLeuAspThrPheIleProAlaArgLysAsnPh 791
841 GGAATATCCAACAGATTAAATGGCTGGCCAAGAGCTCAGCTATAC 886
      :||| ||| ::::: |||||
791 eGluMetProHisAspIleAlaIleAlaGlyAspAspGlyThrValTyr 806

seq_name: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-540-804-12

seq_documentation_block:
; Sequence 12, Application US/08540804
; Patent No. 5919666
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A.
; APPLICANT: Koleske, Anthony J.
; APPLICANT: Thompson, Craigh M.
; APPLICANT: Chao, David M.
; TITLE OF INVENTION: No. 5919666 Factors Which Modify Gene
; TITLE OF INVENTION: Transcription and Methods of Use Therefor
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/540,804
; FILING DATE: 11-OCT-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/521,872
; FILING DATE: 21-AUG-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,265
; FILING DATE: 25-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WH194-03A2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1226 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

```

US-08-540-804-12

alignment_scores:
  Quality: 93.50      Length: 489
  Ratio: 0.403      Gaps: 24
  Percent Similarity: 47.444      Percent Identity: 19.836

alignment_block:
US-09-323-427-1 x US-08-540-804-12 ..

Align seg 1/1 to: US-08-540-804-12 from: 1 to: 1226

513 CATATGAGTACAAATGCTTTTACATGAAAGCTGATTAACAGTTA..... 557
      ::::: ::::: ||| |||||
279 TyrValAspIleAsnProLeuThrMetHisLysIleIleGlnLeuIle 295
558 .....GTGCACAGATTGAG.....TTCGAAATCAGAAC 588
      ::::: ||| ||| ::::: |||
295 uTPSerIleHisProSerArgGlnPheAspHisTyrGlnSerAsnGln 312
589 TGCTTTTCAACTCAAAATTTGCCGATGCCAGTATGCCGTTATGAATTT 638
      ||::: ||||| ||| ||| :::
312 eValAlaLysLeuLeuLeuArgIleAsn..... 322
639 TGGATGATGACCAACCGGTCAACAGTTCAAT.....TTCGT 676
      ::::: ::::: ||| ::::: |||
323 ...SerThrAspLysPheHisGlnPheGlnIleGluAspAlaIleTr 338
677 ATCATTTGGTCAGCAGTTTATCATTAATGACATGCGATTGAAACCGT 726
      ||||| ||| |||
338 pSerLeuValAlaPheGlnLeuAlaLysAsn..... 347
727 TGATATTCTTCTGCGGTTGTCATTCCTTTTCGATGATGTAAG 776
      ::::: ||| ::::: |||
348 .....PheSerAlaGlnLysArgValAlaSer 356
777 GTGATAGTGTGAATTTCAATGATGATGATGCTGTTGATAAAT 826
      ::::: ::::: |||||
357 TyrMetLeuProSerLeuTyrArgLeuLeuAsnIleLeu..... 369
827 TTGCTAATAATTTGGAATATCCACAGATTTAATGCTGCGCCAGAAC 876
      ||| ::::: ||| ::::: |||
370 .....IleThrTyrGlyIleIleLysValProThrTyrIleArgLysL 384
877 TCACGTAT.....ACAAATGCGGATCGAT 902
      ||| ::::: ||| ::::: |||
384 eulIeSerSerGlyLeuLeuTyrLeuGlnAspSerAsnAspLysPheVal 400
903 CAC..AGCTTTTCTATCAATGCCAGATCAGTATTAACATTAAAGAAC 949
      ||| ::::: ||||| ::::: |||
401 HisValGlnLeuLeuIleAsnLeuLysIleSerProLeuMetLysSer 417
950 A.....ATAGCAATGCTGTCACCAACCAATGTCAGAACCA 987
      |::: ||||| ::::: ||| ::::: |||
417 nTyrAsnMetValLeuArgAsnValMetIuTyrAspValLysPheTyrG 434
988 AG.....GATTGCGAGCTGTTAAACAGGTGCGCCAGCAAAACGTG 1031
      ::::: ::::: ||| ::::: |||
434 IuIlePheAsnPheAspGlnLeuValGlnIleThrGlnGlnIleLysMet 450
1032 CTGCAGTCGCGCAACTTCGTTTACTCAAGAAAAGATCTGCAGACCGGAG 1081
      ||| ||| ||| ::::: |||
451 ArgIleLeuSerAsn.....AspIleThrAsnLeu 461
1082 AATATCATTTGATGTACAGACTGATATCAACACCTTGAATTAAGCATGA 1131
      ::::: ||| ::::: |||
461 nLeuSer.....LysThrProLeuSerIleLysIle 472
1132 TATATC.....AAGCTTGGCAGTGTGATTACGTCAAC 1163
      ::::: ||| ::::: |||
472 eValAlaGluTyrTyrLeuSerHisLeuLysSerGlyIleLeuSerSer 488

```

```

1164 GTGCAC.....TTCGCAACATTAAT 1183
      |||||
489 ValAsnArgThrValLeuLeuLysIlePheLysIlePheCysIleAspLe 505
1184 GGACAACCGTATACTTCTGTCAGTACAAATGGAATCTGCATGTCACC 1233
      :::::
505 uGluValPheHisHisPhePheLysTrpIleGluPheIleValTyrHisG 522
1234 ATTGGCTTCTCAATGTTATGGGTTTAAGCATTCATTCATGATGTCGGCG 1283
      ::||| |||||
522 InLeuLeuSerAsp.....IleGluSerLeuGluAlaLeuMetAsp 535
1284 TCATTATTACATTCCTGTTAAATTCGTCAA...ATCAGAGGCGAT.. 1328
      |||||
536 IleLeuLeuCysTyrGlnLysLeuPheSerGlnPheIleAsnAspHisI 552
1329 .....AAAAATATGTTAGATTCATCGAAGCAATAAATAA... 1364
      |||||
552 eleuPheThrLysThrPheIlePheIleTyrLysLysValLeuLysGluL 569
1365 .....CTGCCATATATATTCGTTCTTCTTATCATCCTTCATATACCTAA 1409
      ::||| |||||
569 ysAspValProAlaTyrAsnValThrSerPheMetProPheTrpLys... 584
1410 TTTTACCTTACAAATATATATATAT.....GTAGGAATATATATCTG 1450
      |||
585 PhePheMetLysAsnPheProPheValLeuLysValAspAsnAspLeuAr 601
1451 TAATCAATTAAGATATTTTC.....ATCAAACTCTCTCTCA 1488
      :::::
601 gTleGluLeuGlnSerValTyrAsnAspGluLysLeuLysThrGluLysL 618
1489 TCGCTTTTACCTTCTGAAAAGCTTATTCATTCATTCAGTAATC..... 1532
      ::||| |||||
618 euLysAsnAspLysSerGluValLeuLysValTyrSerMetIleAsnAsn 634
1532 ..... 1532
635 SerAsnGlnAlaValGlyLysInThrTyrAsnPheProGluValPheGlnA 651
1533 .....TTTATATGCATACATTTGTAATGTTTCATCATTTATGGCC 1572
      |||||
651 LAsnIleArgPheLeuLeuHisAsnSerGluIleIleAspThrAsnThrS 668
1573 ATGAATAGTTCGTTGTTTATATATCATTCATTCACCTTGCTCATTTTA 1622
      ::|||
668 erLysGlnPheGlnLysAlaArgAsnAsnVal..... 678
1623 TTCTACACATTATCATTTGTGATTAATATCACAATTAATACCTTGATTG 1672
      :::::
679 .....MetLeuLeuIleAlaThrAsnLeuLysGluTyrIleLysPheK 693
1673 CCCAATTTTATGGGCAATTCCTATTCGTAAACAATTCACCTATATT 1722
      |||||
693 fSerIlePheLeuLysArgLysAspPheThr...AsnLysAsnLeuIleG 709
1723 GCATTATTCGAATTA 1739
      ::|||
709 InLeuIleSerLeuLys 714
seq_name: /cgn2_6/ptodata/1/iaa/5B.COMB.pep:us-08-218-265-12
seq_documentation_block:
: Sequence 12, Application US/08218265
: Patent NO. 5922585
: GENERAL INFORMATION:
: APPLICANT: Young, Richard A.
: APPLICANT: Koleske, Anthony J.
: APPLICANT: Thompson, Craig M.
: TITLE OF INVENTION: No. 5922585el Factors Which Modify Gene
: NUMBER OF SEQUENCES: 35
: TRANSCRIPTION AND METHODS OF USE THEREOF
: CORRESPONDENCE ADDRESS:

```

```

: ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
: STREET: Two Millitia Drive
: CITY: Lexington
: STATE: MA
: COUNTRY: US
: ZIP: 02173
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/218,265
: FILING DATE: 25-MAR-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Granahan, Patricia
: REGISTRATION NUMBER: 32,227
: REFERENCE/DOCKET NUMBER: WHI94-03
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-861-9540
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1226 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-218-265-12

```

```

alignment_scores:
      quality: 93.50      length: 489
      ratio: 0.403      gaps: 24
Percent Similarity: 47.444      Percent Identity: 19.836

```

```

alignment_block:
US-09-323-427-1 x US-08-218-265-12 ..

```

```

Align seg 1/1 to: US-08-218-265-12 from: 1 to: 1226

```

```

513 CATATCGAGTACAAATGCTTTTACATGGAAGCTGATTAACAGTTA..... 557
      ::::: ||| |||||
279 TyrValAspGlnAsnProLeuThrMetHisLysIleIleGlnLeuIleLe 295
558 .....GTGCACAGATGAGG.....TATCGAATTCACAAAC 588
295 uTyrSerIleHisProSerArgGlnPheAspHisTyrGlnSerAsnGlnL 312
589 TGCCTTTTCAAACTCAAAATTTGCCGATGCCAGTATGCGGTATGAATTT 638
      ||::: ||||| ||| |||
312 euValAlaLysLeuLeuLeuLeuArgIleAsn..... 322
639 TGGAGGTGGACCAACCGGTCAACAGTTCAAT.....TTGCT 676
      ::|||
323 ..SerThrAspGluAspLeuHisGlnPheGlnIleGluAspAlaIleTr 338
677 ATCATTCGTGACGACCAAGTTATCATTAATGACATGCGAATTCGAACCGT 726
      |||||
338 pSerLeuValPheGlnLeuAlaLysAsn..... 347
727 TGATACCTTCTGCGGGTTGTCATTCCTGTTGTCGATGATGAGTAAAG 776
      ::|||
348 .....PheSerAlaGlnLysArgValValSer 356
777 GTGATACGTGTGAATTCATAATGCTGATGATGATGCTGCTTGTATAAAT 826
      :::::
357 TyrMetMetProSerLeuTyrArgLeuAsnIleLeu..... 369
827 TTGCTAAATTAATTTGGAATATCCACAGATTAATGCGTGGCCCAAGAAC 876
      |||
370 .....IleThrTyrGlyIleIleLysValProThrTyrIleArgLysL 384

```

```

877 TCACGATAT.....ACAAATATGCGGATCGAT 902
   || .....
384 eu1eserserGlyLeuLeuTyLeuGlnAspSerAsnAspLysPheVal 400
903 CAC...AGCTTTTCATCATGCGCAATGATATTCATTTAAGAACCA 949
   || .....
401 HisValGlnLeuLeuLeuLeuLysLieserProLeuMetLysSerGI 417
950 A.....ATAGCGATGTGTTCGACGACCAATGTCAGAACCA 987
   || .....
417 nTyAsnMetValLeuArgAsnValMetGluTyAspValLysPheTyRG 434
988 AG.....CATTCGAGCTGTAAACAGCTGTGCGCCAGCAAAACCTG 1031
   || .....
434 InLePheAsnPheAspGlnLeuValGlnLePheGlnGlnLeLysMet 450
1032 CTGCAAGCTGCGACCTGCTTACTCAAGAAAGATCTCAGACCGGAG 1081
   || .....
451 ArgLLeuSerAsn.....AspLLePheAsnLeuGI 461
1082 AATATCATGTATGATAGCAATGATATCAACACCCCTGAATTAGCGATGA 1131
   || .....
461 nLeuSer.....LysThrProLeuSerLLeLysLLeM 472
1132 TAATC.....AAGCTTGGCAGTTGATTTAGCTGACC 1163
   || .....
472 euValAlaGlnTrpTyLeuSerHisLeuCysSerGlyLeuLeuSer 488
1164 GTGCAAC.....TCTGCAACATTAAT 1183
   || .....
489 ValAsnArgThValLeuLeuLysLLePheLysLLePheCysLLeAspLe 505
1184 GGACAACCTGTATAGCTCTGCAAGTACAAATGGAATCTGCATGTCACC 1233
   || .....
505 uGlnValPheHisPhePheLeuStrPLeGlnPheLLeValTyHisG 522
1234 ATTGGCTCTCAATGTTTATGGGTTTAAGCATGTCATGATTCGTCGGC 1283
   || .....
522 InLeuLeuSerAsp.....LeGlnSerLeuGlnAlaLeuMetAsp 535
1284 TCATATTATGCAATTCGTTAAATTCGTCGCAAA...ATCAGAAGCAT... 1328
   || .....
536 LLeLeuLeuCysTyGlnLysLeuPheSerGlnPheLLeAsnAspHisL 552
1329 .....AAAAATATGTTAGATTCATGCAAGCAATGAATAAAA... 1364
   || .....
552 euLeuPheTrpLysTrnPheLLePheLLeTyLysLysValLeuLysGln 569
1365 .....CTGCCATATATTCGCTTCTCTATCATCCCTCTAATAACTAA 1409
   || .....
569 yAspValPheAlaTyAsnValTrpSerPheMetProPheTrpLys... 584
1410 TTTTGGCTAACAATATATATAGTAT.....GTAGAAATATATAGTATG 1450
   || .....
585 PhePheMetLysAsnPheProPheValLeuLysValAspAsnAspLeuArg 601
1451 TAATACAAATAGATATTTT.....ATCAAAACTTCTCTA 1488
   || .....
601 gLLeGlnLeuGlnSerValTyAsnAspGlnTyLysLeuLysTrpGlnLysL 618
1489 TGGCTTTTATGCTCTGAAAGCTTATGATTTAGTATC..... 1532
   || .....
618 euLysAsnAspLysSerGlnValLeuLysValTyLysMetLLeAsnAsn 634
1532 ..... 1532
635 SerAsnGlnAlaValGlnTrpTrpAsnPheProGlnValPheGlnVal 651
1533 .....TTTATATGCAATCTATGTTAATGTTTCATCATTAAGGCC 1572
   || .....
651 LAsnLLeArgPheLeuLeuHisAsnSerGlnLLeLLeAspTrpHis 668
1573 ATGAATAGTGTGTTGTTATATGATCATGTATGCAACTGTGCTATTTTA 1622

```

```

668 euLysGlnPheGlnLysAlaArgAsnAsnVal..... 678
1623 TTTCTAAGCTTATCATTTGATGATATATACAAATATATACCTTGATTCG 1672
   || .....
679 .....MetLeuLeuLLeAlaThrAsnLeuLysGlnTyTrLLeLysPheM 693
1673 CCCAATTTTATGGGCGATCATTTCCATATTCGTAAACAAATCATCTATTT 1722
   || .....
693 tSerLLePheLLeLysArgLysAspPheThr...AsnLysAsnLeuLLeG 709
1723 GCATTTATGCAATTA 1739
   || .....
709 InLeuLLeSerLeuLys 714

```

seq_name: /cgn2_6/ptodata/1/laa/5A_COMB pep:US-07-906-349A-6

```

seq_documentation_block:
: Sequence 6, Application US/07906349A
: Patent No. 5434064
: GENERAL INFORMATION:
: APPLICANT: Schlessinger, Joseph
: APPLICANT: Skolnik, Edward Y.
: APPLICANT: Margolis, Benjamin L.
: TITLE OF INVENTION: A NOVEL EXPRESSION-CLONING METHOD FOR
: TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE KINASES
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Browdy and Neimark
: STREET: 419 Seventh Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent in Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/906,349A
: FILING DATE: 30-JUN-1992
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/643,237
: FILING DATE: 18-JAN-1991
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-628-5197
: TELEFAX: 202-737-3528
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 801 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-07-906-349A-6

```

```

alignment_scores:
  Quality: 93.00      Length: 77
  Ratio: 3.000        Gaps: 1
  Percent Similarity: 40.260  Percent Identity: 28.571

```

alignment_block:
US-09-323-427-1 x US-07-906-349A-6 ..

Align seg 1/1 to: US-07-906-349A-6 from: 1 to: 801

```

982 ACCACAGAGATTGCGAGCTGTTAAACAGTGTGCGCCAGCAAAACCTG 1031
   ||||| .....
573 ThrTrpAlaAlaIleThrCysAlaCysThrGlyCysTrpHisCys 589

```



```

1068 ...TCGGAAGACCGGAAATTCATTTGATGACAGCATGATTCACA 1111
      ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1119 euValLeuValAsnHisLysIleSerLysIleHisGlyLysAlaPheSer 135
      |||||::: :::: ||| ::|||:::
136 ProLeuArgAsnVal .....GlnLysLeuTyIleSe 146
1163 CGTGCACCTTCTGCAACATATATGCAACAACCTGTAACTTGTGCGAGTCA 1212
      :
146 rLys..... 147
1213 AATGGAATCTGCATGTCACCACTTTGGCTCTGCAATGTTATAGGGTTAA 1262
      ::||| |||||:::
148 .....AsnHisLeuValIu..... 152
1263 GCATTGCAATTGATTCGCTGCCGTCATTAATTCACATTTGCTTAAATTTGCT 1312
      ::||| ||||| ||| ||
153 .....IleProAsnLeuProSerSerLeu.....Va 162
1313 CCAATACAGACAGCGCATTAATAATATGTTAA.....ATCA 1347
      |::: |||| ||| |||||::: ||
162 AspValArgIleHisAspAsnArgIleArgLysValProLysGlyValR 179
      |::: |||| ||| |||||::: ||
1348 TCGAAGACATA..... 1358
      ::|||:::
179 heserLLeuLeuArgAsnMetAsnCysIleGlnMetGlyGlnAspProLeu 195
      ::|||:::
1359 .....ATAAACTGCCATA 1372
      ::||| ||| ||
196 GlnAsnSerGlyPheGlyProGlyAlaPheAspGlyLeuLysLeuAsnTy 212
      |::: |||||::: |||||
1373 TAAATTCGTTTCTGTCATATACCTCTGTATACATACTATTTAGTACASCA 1422
      |::: |||||::: |||||
212 rLeuArgIleSerGlnAlaLys.....LeuThrG 222
      |::: |||||::: |||||
1423 ATATATAGTATGATGAGAAATATTAAGTATATACAAATAGAGATATTTC 1472
      ::||| ::|||::: ::|||:::
222 TyIle.....ProLysAspLeuProLyrHisLeuAsnGlnLeuHis 235
1473 ATCAAAATCTTCTCTATGCGCTTTTATAGCTTCTGAAAAGCTTATTCATA 1522
      ::||| ::|||::: ||| |||
236 LeuAspHisAsnLysIleGlnAlaIleGlnLeuGlnSerLeuLeuArgTy 252
      |::: |||||::: |||||
1523 TCACAGTAATCTTATATGCAATGATATGTAATGTTCAATCATTAAGCC 1572
      ||||| ::|||::: ::|||:::
252 rSerLysLeuTyArgLeu.....GlyLeuGln 262
      |::: |||||::: |||||
1573 ATGCAATAGTTTCTGTTTATATATCA.....TCATTAATCAACTGTCT 1616
      ||::: |||||::: ::|||:::
262 IAsnGlnIleLeuArgMetIleGlnAsnGlySerLeuSerPheLeuProThr 278
1616 ..... 1616
279 LeuArgGlnLeuHisLeuAspAsnLysLeuAlaArgValProSerGln 295
1617 .....ATTATATCTCAACAGTTTATATGATTTGTGATATA 1650
      ::||| ::|||::: ::|||:::
295 yLeuProAspLeuLysLeuLeuGlnValValTyuLeuHisSerLysAsnI 312
1651 TCACAAATTAATACCTTG.....TATGCGCAATTTTATAGGAGCATC 1691
      |||||::: ::|||::: |||||:::
312 LeHisLysValGlyValAsnAspPheCysTrpMetGlyPheGlnVal 327
seq_name: /cgn2_6/ptodate/1/aa/5A.COMB.rep:05-08-325-071-67

```

```
seq_documentation_block:
; Sequence 67, Application US/08325071
```

```

;
; GENERAL INFORMATION:
;
; APPLICANT: COBON, Stewart Gary
;
; APPLICANT: MOORE, Joanna Terry

```

APPLICANT: JOHNSON, Law Anthony York
 APPLICANT: WILLADSEN, Peter
 APPLICANT: KEMP, David Harold
 APPLICANT: SRINANTHA, Alagaccone
 APPLICANT: RIDING, George Alfred
 APPLICANT: RAND, Keith No. 5587311man
 TITLE OF INVENTION: DNA Encoding A Cell Membrane
 TITLE OF INVENTION: Glycoprotein Of A Tick Gut
 NUMBER OF SEQUENCES: 71
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 3000 K Street, N.W.
 CITY: Washington, D.C.
 COUNTRY: USA
 ZIP: 20007-5109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/325,071
 FILING DATE: 14-OCT-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/062,109
 FILING DATE: 17-MAY-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/926,368
 FILING DATE: 07-AUG-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/242,196
 FILING DATE: 06-JUL-1988
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/AU87/00401
 FILING DATE: 27-NOV-1987
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: AU P14912
 FILING DATE: 16-OCT-1987
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: AU P12570
 FILING DATE: 19-JUN-1987
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: AU P49196
 FILING DATE: 27-NOV-1986
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 60042/111 BIAU
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202 672 5300
 TELEFAX: 202 672 5399
 TELEX: 904136
 INFORMATION FOR SEQ ID NO: 67:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 650 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

```
alignment_scores:
  Quality: 90.00
  Ratio: 0.612
  Gaps: 22
Percent Similarity: 40.947
Percent Identity: 21.727
```

```
alignment_block:
```

US-09-323-427-1 X US-08-325-071-67

Align seg 1/1 to: US-08-325-071-67 from: 1 to: 650

304 TGCATTGAGGACATGT.....TTATG 326
 |||::: ||||| |||||

```

147 CysValProThrThrcysLeuArgProaspLeuThrCysLysaspLeu 163
327 TGAAGAGTCTTATGATCAAGAGTTGCGGTATGATGAGTGACGG 376
163 sglulysleuleuLeuGlnArgasp.....SerA 173
377 CAAGTTGCCGGAATTCACCTTCATTTGATTCATGCAATGTCGGGTAC 426
173 rgycysgslnglyrPasnThr.....AlasnCysSerAlaIa 186
427 AGCATCTCGAATCCACGTGTATTTTGTAC..... 459
187 ProProAlaaspser.....TyrCysSerProGlySerProLysgl 200
460 .....ACAACTGTGTGCAT.....TTCG 478
200 yProaspGlyGlnCysLysAsnAlaCysArgThrLysGlnAlaGlyPhe 217
479 TTTCATCATTTATTTGTTACCAAGTTGATCGTATTCGATTCGATACAG 528
217 alCysLysHsLysGlyCys.ArgSerThrAspLysAlaLysGlnCysThrcy 233
529 C.....TTTACATGAGACCTGATAAACAGTTAGTGCACAGATTG 569
233 sProSerGlySerThrValAlaGlnLysGlyLethrCysLysSerIles 250
570 AGGATCTCTGAATCACACAGCTGTTTCAAACTCAATTC.....CCGATG 616
250 ertyrThrValSerCysThrValGlnLysGlnLysThrCysArgProThr 266
617 CCAGTATGCCGTTATGAA.....ATTGATGATGGGACCAACCGG 657
267 GlnAspCysArgValGlnLysGlyThrValLeuCysGlnCysProTrpAs 283
658 TCACACAGATTCAATTGCTATCTGTCAGCCAGCTTATTCATTAATGGA 707
283 ngln.....HisLeuValGlyasp.....T 290
708 CATGCGATTCTGAACCGTTGATACCTTTGCGCGGTGTCATTCCTGC 757
290 hCysIleSeraspCysValaspLysLysCys.....HisGlnGlu 303
758 TTGTGTGATGATGCT..... 772
304 PheMetAspCysGlyValAlaIleAsnArgGlnSerCysIleCysProTr 320
773 .....AACGTGATCTGGAATATTCAATGCTGATGAT 809
320 pLysSerArgLysProGlyProAsnVal.....AsnIleAsnGlnC 334
810 GTGCTCTGTATAAATATTGCTAAAT.....AATTTGGA 844
334 yLeuLeuAsnGlnIleGlyrThrValSerPheThrProAsnIleSer 350
845 TATCCACAGATTATGCTGCGCCAAAGACTCAGTATACAAATATGC 894
351 PheAspSerasp.....HisCysLysArgTyrGln 360
895 GGATCGATACAGCTTTTCTATCAATGCCAGATCACTATTACCATTAAG 944
360 unAspArg.....ValLeuGlnAlaIleArgThrSerIleGlyLysGln 375
945 AACCAAAATAGCGAATGTTCGACCACAATGTTCAGAACCAAGAGATTG 994
375 alPheLysValGlnIleLeu.....AsnCysThrGlnasp..... 386
995 GGAAGCTTAAAAACAGTGTGCGCCGACGAAAAACCTGTCGACACTCGCA 1044
387 .....IleLysAlaArgLeuIleAlaGlnLysProLeuSerLysTyrVa 401
1045 ACTTCT..... 1051
401 lleuArgLysLeuGlnAlaCysGlnHisProIleGlyGlnTrpCysMet 418

```

```

1052 .....TTCACAGAAAAGATCTGA.....GAACCG 1078
418 ertyrProLysLeuLeuIleLysAsnSerAlaThrLuiIleGlnGlu 434
1079 GAGATATATGATGATGTACCAACTGATATCAACACCCCTGGAATTCGCA 1128
435 GlnAsnLeuCysAspSerLeuLeuLysAsnGlnAlaAlaLysLysGln 451
1129 TGATATCAAGCTTGCACGATGAT 1153
451 yGlnAsnLysCysValLysValasp 459
seq_name: /cgn2_6/ptodate/1/1aa/58_COMB.pep:US-08-477-451-22

```

```

seq_documentation_block:
; Sequence 22, Application US/08477451
; Patent No. 5928865
; GENERAL INFORMATION:
; APPLICANT: Covacc, Antonello
; TITLE OF INVENTION: Helicobacter Pylori CagI Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESS: Chilton Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,451
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0335,002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2708
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 485 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-477-451-22

```

```

alignment_scores:
  Quality: 89.50      Length: 564
  Ratio: 0.393      Gaps: 28
  Percent Similarity: 40.426      Percent Identity: 18.085

```

```

alignment_block:
US-09-323-427-1 x US-08-477-451-22 ..

```

```

Align seg 1/1 to: US-08-477-451-22 from: 1 to: 485
333 GTCCTTATGATCAAGAGTTCGCCGTAATGATGAAGGTGGATCAAGTT 382
10 ValIleLeuValTyrProIleLeuPheLeuPheAlaLeuIleLys.. 25
383 GCCGGAATTCACCTTCATTTGATTCATGCAATGTTGCGCGCTA..... 425
26 .ProSerPhePheTyrTyrThrThrTyrIleuLeuLeuValSerLeu 42

```

```

426 .....CAGCATCTGATCCAGCTG..... 446
42 erlleleSerlystrYrleuLeuSerHisAlaLysPheThrMetLys 58
447 .....GATTTTGGACA 460
59 leulleleleuMetThrGlnTrpGlnsntPrpheLeuPheValGlnly 75
461 ACAAGTGTTCATTTGTTTCATCCATTAT.....TTGT 495
75 spherleuLysThrPheArgGlnleSerLeuPheProAsnleSerleL 92
456 TACCAAGTGTATCGTCATATCGATACAAATGCTTTTACATGGAAGCTG 545
92 ySPrOTyLeuLysThrLyslePheLeuSerAlaYrThr..... 105
546 ATAAAAAGATTAGTGACAGATTGAGTATCGAATACACACTGCTTT 595
106 .....LysleuAlaYrleuLysLysProileYrPh 116
596 C.....AACTCAAAATTGT 609
116 eTrpYrSerAsnProLysThrProPheleValSerValPheLysPheC 133
610 CCCGATGCCAGATGCCGTTATGAATTTGATGCGTGCACCAACCGG.. 657
133 yslleCysValSerAsnLeuPheleuileTrp.TrpTrpPheSerLysle 149
658 ..TCAACCAATTCAATTTGCTATCAT..... 681
149 uilleleuileLysleCysYrAsnGlnleLysThrLysArgileTrpL 166
682 .....TGTCAGCCAGTTTATCATAAATGACATGCGAT..... 715
166 euValTrpTrp.GlnAsnlePhe.....ValCysAspTrpTrple 179
716 .....TCGAAACGCTGATACTTTCTGCGGCTTG 746
179 uleuileAlaPheHisGlnSerYrGlnleGlnGlnPheSerleuVal.. 195
747 TCCATTCCTGCTTT.....GTGCGATGAT 769
196 .....CysTyrCysGlnMetleuProThrLysThrArgileAlaGAsp 209
770 GGTAACGCGTACTCTGGAATTTCTAAATGCTGATGATGCTGCTTGA 819
210 ProAsnLysGlnGlnleuThrGlnProLysleLysGlyleuileMetG1 226
820 TTAATrTTTGTAAATATTTGGAATATCCACACAGATTAAATGCGTGGC 869
226 yLysleleuAlaSerleuLeuGlyGlyYrAsnleuPheThrGlyL 243
870 AAGAAGCTCAGTATACAAATATGCGATGATGATGATGATGATGATGAT 911
243 euSerSerAspLeuPhe..SerMetleuAsnPheLeuPhePheleuNe 259
912 .....TCTATCAATGCCAGATCAGATTATACA. 938
259 tleuMetGlyleuAsnGlnAlaLeuGlyLysLysPheAsnleuProm 276
939 .....TTAAGAACCAATAGCGAATGTGTGACACACAAATGTTGAG 980
276 eTAspAsnleLysAsnPheMetAlaGlnValleuLysAsnGlyPheAsp 292
981 AACCAACAG..... 989
293 SerleLysAsnMetGlySerAlaLeuValGlyAsnGlyPheGlySerSe 309
990 .....GATTCGAGAGCTGTAAACAGGTGTCGCCGACGCAA..... 1025
309 rlySerAspLysThrAlaAsnLysMetSerValSerGlnValArgleuTr 326
1026 .....AACCTGCTGCAGCTGCCCACTTCTTACTCAAGAAAGAT 1067

```

```

326 rPCyslleAsnYrleuLysValLysPhePheHisLeuSerTrgArgAla 342
1068 CTG.....CAGAACCGAGAAATATATCATTTGATGATACGATGAT 1105
343 leuPheAspGlnArgGlyAsnArgleuLeuAlaGlnHisTyrAspPhe1 359
1106 ATCAACACCC...TTGAATTTAGCATGATAATCAAGCTTTGCCAGTTGA 1152
359 eThrAspSerPheleuLysMetSerYrleuAspProleuPheLeuArg1 376
1153 TTTAGCTCACCC...GTGCACTTCTGCACATATATGACACACTGTATATAC 1199
376 lePheThrThrAspLeuAlaPheCysGln..... 385
1200 TTTGCTCAGTACAAATGGAATGTCATGTCACCACTTGGCTTCAATG 1249
385 ..... 385
1250 TTTATGGTTTAAAGCATTTGCATTGATGCTGCGGTCATTATTACCATTTG 1299
386 .....LeuProThrleuLys.....A 391
1300 GTTTAAATTTGCTCCAAATCAGAGGCAATAAATATATGTTAGATCATC 1349
391 rgleuGlyPhePheleuHisArg..... 399
1350 GAAGCAATATATAAACTGCCATATATATGTTCTTCTTATCATTCCTTC 1399
400 ...SerleThrSerleGlnYrValAlaAlaPhePheArgAsnAsnPro1 415
1400 TATATACTATATTTTACTATACAAATATATATATATATATATATACT 1449
415 eArgGly.....S 418
1450 GTAATATATATAGTATATTTTATCATCAAACTTCTTATGCTTTTATA 1499
418 eSerThrSerPheAspAlaLeuileSerValMetSerVal.....lle 432
1500 GCTTTCGAAGCTTATTCATTATTCAGTAATCTTTTATATGCAATACATAT 1549
433 SerSerAspAsnProLysrProThrPheleYrleuYrPheMetProva 449
1550 TGTAAATGTTTCATCATTTAGGCCATGATAGTTTCGTTTGTATATATCAT 1599
449 lleHisAlaSerValValSerAsnleAspAlaArgileValilleSer. 465
1600 CATATACACTTGTCTATTTATTTCTACAGTTTATCATTTTGTGTATAT 1649
466 .....ValValCysAspIle 470
1650 ATCACAATATATACCTTGTATTTGCCCAATTTTATG 1685
471 AlaMetArgThrSerGlnPheSerAspValPheMet 482

```

```

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB pep: US-08-325-071-61
seq_documentation_block:
: Sequence 61, Application us/08325071
: Patent No. 5587311
: GENERAL INFORMATION:
: APPLICANT: COBON, Stewart Gary
: APPLICANT: MOORE, Joanna Terry
: APPLICANT: JOHNSON, Law Anthony York
: APPLICANT: WILADSEN, Peter
: APPLICANT: KEMP, David Harold
: APPLICANT: SRISKANTHA, Alagagone
: APPLICANT: RIDING, George Alfred
: TITLE OF INVENTION: DNA Encoding A Human
: TITLE OF INVENTION: DNA Encoding A Cell Membrane
: NUMBER OF SEQUENCES: 71
: CORRESPONDENCE ADDRESS:

```

ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/325,071
FILING DATE: 14-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,109
FILING DATE: 17-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/926,368
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/242,196
FILING DATE: 06-JUL-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU87/00401
FILING DATE: 27-NOV-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P14912
FILING DATE: 16-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P12570
FILING DATE: 19-JUN-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PH9196
FILING DATE: 27-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 60042/111 BIAU
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 549 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-325-071-61
alignment_scores:
Quality: 89.00 Length: 359
Ratio: 0.614 Gaps: 22
Percent Similarity: 40.390 Percent Identity: 22.284
alignment_block:
US-09-323-427-1 x US-08-325-071-61 ..
Align seg 1/1 to: US-08-325-071-61 from: 1 to: 549
304 TGCATTGGAAGACATGT.....TTATG 326
||||| |||||
66 CysValProThrThrCysLeuArgProAspLeuThrCysLysAspLeu 82
|||:||||| :|||
327 TGAAGCTCTTATGATCAAGAAGGTTGCCGTATATGATGACAGTGCACGT 376
|||||:||||| :|||
82 sGluLysAsnLeuLeuGlnArgAsp.....SerA 92
|||:||||| :|||
377 CAAGTGGCGAATTCACCTTCATTTGATTCATGCAATGTCGCGGTAC 426
|||||:||||| :|||
92 rGcYcYcSgInGlyTrpAsnThr.....AlaAsnYcSerAlaAla 105

427 ACGATCTGATCATCACAGTGGTATTTTGTAAAC..... 462
:|||||
106 ProProAlaAspSer.....TyrCysSerProGlySerProLysG 119
463ACTGTTGTCAT.....TTGG 478
|||||:|||||
119 yProAspGlyGlnCysIleAsnAlaCysLysMetLysGlnAlaGlyPhe 136
479 TTTCATCATATTATTGTTTACCAAAAGTTGATCGCATATGATGATACAAAG 528
|||:||||| :|||
136 alCysGlnHisGlyCys.ArgSerThrAlaLysAlaTyrGluCysThrC 152
529 C.....TTTACATGGAACCTGATATAAACACGTTAGTCACAGATTG 569
| ||| :|||
152 sProArgGlyPheThrValAlaGlnAspGlyIleThrCysLysSerIle 169
470 AGGTATCTGAAATCACAACTGCTTTTCAAACTCAATTGTC..CCGAG 616
: :||| ||||| ||| |||
169 eRhISThrValSerCysThrAlaGlnLysGlnThrCysArgProThr 185
617 CCAGTATGCGGTATGAA.....ATTGTGATGGTGGACCAACCGG 657
||||| :|||
186 GluAspCysArgValHisLysGlyThrValLeuCysGluCysProTrpAs 202
||||| :|||
202 nGln.....HisLeuValGlyAsp.....T 209
708 CATGCGATTCTGAACCGTTGATACTTCTCGCGGTTGTCATTCTCTGC 757
||||| :||||| |||||
209 hrcYsIleSerAspCysValAspLysLysCys.....HisGluGlu 222
758 TTGTGATGATGATGT..... 772
|||||:|||||
223 PheMetAspCysGlyValAlaTyrMetAsnArgGlnSerCysTyrCysProTr 239
239 pLysSerArgLysProGlyProAsnVal.....AsnIleAsnGlyC 253
810 GTGCTCTTGATAATATTGCTTAAT.....ATTGGAA 844
|||:||||| :|||
253 ySLeuLeuAsnGlnTyrTyrThrValSerPheThrProAsnIleSer 269
845 TATCAACAGATTTAATGCTGCGCAAGAAGCTCACGTATACAAATATGC 894
: :||| |||
270 PheAspSerAsp.....HisCysLysTrpTrpG 279
895 GCATGCATCACAGCTTTTCTATCAATGCCAGATCAGTATTACCATTAAG 944
||||| :|||
279 uAspArg.....ValLeuGlnAlaIleArgThrSerIleGlyLysGlu 294
945 AACCAATAGCGAATGTGTTCGACCAACATGTTCAACACACAGAGATTG 994
: :||| :|||
294 alPheLysValGluIleLeu.....AsnCysThrGlnAsp..... 305
995 GGAGCTGTTAAACACAGTGTGCGCGACGAAACCTGTCGACGTCGCGA 1044
:|||||:|||||
306IleLysAlaArgLeuIleAlaGluLysProLeuSerAsnHisVa 320
1045 ACTTGCT..... 1051
|||||
320 lleuArgLysLeuGlnAlaCysGlnHisProIleGlyGluTrpCysMet 337
1052TTACTCAGAAAGATTCGCA.....GACCG 1078
|||||:|||||:|||||
337 eTyrProLysLeuLeuIleLysLysAsnSerAlaThrGluIleGluGlu 353
1079 GAGATATCATGATGTATGACACTGATATCAACACCGCTGAATATGACGA 1128
|||||:||||| :|||
354 GluAsnLeuCysAspSerLeuLeuLysAsnGlnGluAlaIleTyrLysG 370
1129 TGATAATCAAGCTTTGCGAGTTGAT 1153

seq_name: /cgn2_6/plodata/1/iaa/5A_COMB.pep:us-08-325-071-63

seq_documentation_block:

; Sequence 63, Application US/08325071

; Patent No. 5587311

; GENERAL INFORMATION:

; APPLICANT: COBON, Stewart Gary

; APPLICANT: MOORE, Joanna Terry

; APPLICANT: JOHNSON, Law Anthony York

; APPLICANT: WILLADSEN, Peter

; APPLICANT: KEMP, David Harold

; APPLICANT: SRISKANTHA, Alagacone

; APPLICANT: RIDING, George Alfred

; APPLICANT: RAND, Keith No. 5587311man

; TITLE OF INVENTION: DNA Encoding A Cell Membrane

; TITLE OF INVENTION: Glycoprotein of A Tick Gut

; NUMBER OF SEQUENCES: 71

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W.

; CITY: Washington, D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/325,071

; FILING DATE: 14-OCT-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/062,109

; FILING DATE: 17-MAY-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/926,368

; FILING DATE: 07-AUG-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/242,196

; FILING DATE: 06-JUL-1988

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/AUS87/00401

; FILING DATE: 27-NOV-1987

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: AU P14912

; FILING DATE: 16-OCT-1987

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: AU P12570

; FILING DATE: 19-JUN-1987

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: AU PH9196

; FILING DATE: 27-NOV-1986

; ATTORNEY/AGENT INFORMATION:

; NAME: BENT, Stephen A.

; REGISTRATION NUMBER: 29,768

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202 672 5300

; TELEFAX: 202 672 5399

; TELETYPE: 904136

; INFORMATION FOR SEQ ID NO: 63:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 650 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-325-071-63

alignment_scores:

Quality: 89.00 Length: 359
Ratio: 0.614 Gaps: 22
Percent Similarity: 40.390 Percent Identity: 22.284

alignment_block:

US-09-323-427-1 x US-08-325-071-63

Align seg 1/1 to: US-08-325-071-63 from: 1 to: 650

304 TGCATTCGAGACGATGT.....TTAG 326
||||| |||||
147 CysValProThrThrcysLeuArgProAspLeuThrcysLysAspLeu 163
327 TGAAGGCTTTATGATCAAGAGTTGCCGTAAAGATGAAGTGAGCT 376
||||| |||||
163 sglLysAsnLeuLeuGlnArgAsp.....Sera 173
377 CAAGTTGCCGAATTTCACTTCATTTGATTCAGATGCAATGTCGGTAC 426
||||| |||||
173 rgcyscysglnGlyTrpAsnThr.....AlasncysSerAlaAla 186
427 ACGATCTCGATCCACGCTGATTTTGTACACAC..... 462
||||| |||||
187 ProProAlaAspSer.....TyrCysSerProGlySerProLysG1 200
463ACGTGTGTCAT.....TTGC 478
||||| |||||
200 yProAspGlyGlnCysIleAsnAlacysLysMetLysGluAlaGlyPhe 217
479 TTTATCCATTAATTTGTTTACCAAGTTGATCGCATATGATGATCAATG 528
|| ||| ||| ||||| |||||
217 alcysgluhisglnCys.ArgSerThrAlaLysAlaTyrGlnCysThr 233
529 C.....TTTACATGGAAGCTGATAAACAGTTATGACAGATG 569
|| ||| ||| ||| ||| |||
233 sProArgGlyPheThrValAlaGlnAspGlyIleThrCysLysSerIle 250
570 ACGTATCGAAATCACACACTGCTTTTCAACCTCAATTTGTC...CCGAG 616
|| ||| ||||| ||||| |||||
250 erHisThrValSerCysThrAlaGlnGlnLysGlnThrCysArgProThr 266
267 GluAspCysArgValHisLysGlyThrValLeuCysGlnCysProThr 283
617 CGAGTATGCCGTTATGAA.....ATTGGATGTGTGACCAACCG 657
||||| ||||| ||||| |||||
267 GluAspCysArgValHisLysGlyThrValLeuCysGlnCysProThr 283
658 TCAACCACTTCAATTTGCTATCATTTGTCACGCCAGTTTATCATTAATGA 707
||| ||||| ||||| |||||
283 nGln.....HisLeuValGlyAsp.....T 290
708 CATGCGATTCGAAACCGTTGATTAATTTCTCGCGGTTGTCATTCCTGC 757
||||| ||||| ||||| |||||
290 hrcysIleSerAspCysValAspLysLysCys.....HisGlnGlu 303
758 TTTGTGATGATGATGT..... 772
||||| ||||| ||||| |||||
304 PheMetAspCysGlyValTyrMetAsnArgLysCysTyrCysProThr 320
773AACGTCGATACGTGGAATTTCAATGCTGATGAT 809
320 pLysSerArgLysProGlyProAsnVal.....AsnIleAsnGlyC 334
810 GTGCTTTGATTAATTAATTTGCTAAAT.....AATTGGA 844
|| ||||| ||||| ||||| |||||
334 yLeuLeuAsnGlnTyrTyrTyrThrValSerPheThrProAsnIleSer 350
845 TATCCACACAGATTTAATGCTGCGCAGAGCTCACGTTATACAAATATGC 894
||| ||||| ||||| |||||
351 PheAspSerAsp.....HisCysLysTyrTrpG1 360
895 GGATGCATCACAGCTTTTCTATCAATGCCAGATCAGTATTACATTAAAG 944
||||| ||||| ||||| |||||
360 uAspArg.....ValLeuGluAlaIleArgThrSerIleGlyLysGlu 975

```

945 AACCAATAGCGAATGTGTCGACCAATGTTCAGAACCAAGGATTG 994
    :: ||| :: ::|||:::
375 aPheLysValGIuILeU.....AsnCysThrGlnasp..... 386
    ::|||:::
995 GGAGCTGTAAACAGGTGTGCGCCGACCAAACTGCTGCAGCTGC 1044
    ::|||:::
387 .....ILeYsAlaArgLeuILeAlaGIuLysProLeuSerAsnHis 401
    ::|||:::
1045 ACTTGT..... 1051
    |||||
401 lLeuAlrGLysLeuGlnAlaCysGIuHisProILeGIuTrpCysMet 418
    |||||
1052 .....TTACTCAAGAAAGATCTCA.....GAACG 1078
    |||||
418 eTyrProLysLeuLeuILeLysAsnSerAlaThrGIuILeGIu 434
    |||||
1079 GAGAAATATCATGTGTACGAAGTATATCAACACCCCTGAATTA 1128
    |||||
435 GIuAsnLeuCysAspSerLeuLysAsnGlnGluAlaAlaTyrLys 451
    ::|||
1129 TGATATCAAGCTTTCAGTTGAT 1153
    ::|||
451 yGlnAsnLysCysValLysValasp 459

```


OM of: US-09-323-427-1 to: PIR_62: * out_format : pfs
Date: Apr 16, 2000 4:50 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=frame+azp.model -DEV=xlp
-O=/cgn2.1/uspro.spool/US09323427/runat_14042000_170513_19903/app_query.fasta.1
-DB=PIR_62 -QPM=fastan -SUFFIX=backtrans.fpr -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.500
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DEGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DEGAPEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human0.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=1000000 -USER=US09323427 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
-THREADS=1

Search information block:

Query: US-09-323-427-1
Query length: 1779
Database: PIR_62:
Database sequences: 142080
Database length: 47169319
Search time (sec): 178.310000

score_list:

Sequence	Strd Orig	ZScore	Score len	Documentation
pir2:A49772	1379.00	2.271.64	423	cuticle protein cut-1 - Caenorh
pir2:S27799	897.50	1475.17	308	cuticle protein cut-1 - Caenorh
pir2:S72284	119.00	171.60	0.0093	DNA-directed RNA polymerase (EC
pir2:F71606	116.50	165.96	0.0164	1122 i probable multiple transmembran
pir2:B26696	112.00	167.39	0.0345	443 i hypothetical protein (CYP-CO1
pir1:RN008F	112.00	159.36	0.0418	1024 i DNA-directed RNA polymerase (H
pir2:S78177	108.00	156.14	0.0904	717 i hypothetical protein 717 - Rec1
pir2:A40970	108.00	154.59	0.0938	843 i undulin 1 - human (fragment)
pir2:C71607	107.00	148.40	0.1293	1351 i hypothetical protein PFB0745w
pir2:C71613	105.50	145.95	0.1778	1346 i hypothetical protein PFB0740w
pir2:S55098	104.50	150.00	0.1919	742 i probable membrane protein YMR21
pir2:C71618	104.50	141.98	0.2325	1712 i hypothetical protein PFB0315w
pir2:E71604	104.50	138.82	0.2508	6830 i hypothetical protein PFB0870C
pir2:S57242	104.00	127.88	0.3554	6839 i twitchin - Caenorhabditis eleg
pir2:T03099	104.00	121.51	0.4139	13288 i myosin, submaxillary - pig
pir2:S52967	103.00	150.31	0.2470	554 i NADH dehydrogenase (ubiquinone)
pir2:S72278	103.00	147.21	0.2660	765 i ATP-dependent Clp protease (H
pir2:D71606	103.00	138.92	0.3244	1817 i hypothetical protein PFB0800C
pir2:T15789	103.00	148.92	0.4536	7829 i hypothetical protein C1A3.1 -
pir1:A35269	102.50	147.02	0.2915	716 i translation initiation factor I
pir2:A36054	100.50	146.00	0.4223	563 i myosin homolog - bovine
pir2:S05356	100.00	150.68	0.4117	317 i hypothetical protein (clone pLk
pir2:S17855	100.00	140.31	0.5277	935 i peptidylglycine monooxygenase (
pir1:VGBE11	99.50	142.02	0.5523	717 i glycoprotein H precursor - sain
pir2:E71615	99.50	137.10	0.6140	1099 i S-layer protein precursor - Ba
pir2:T14850	99.00	137.09	0.6776	1100 i SERA antigen/papain-like prote
pir2:H71616	99.00	124.78	0.9039	3973 i hypothetical protein PFB0555C
pir2:C70126	98.50	141.91	0.6585	610 i DNA mismatch repair protein (mu
pir2:T14594	98.50	134.46	0.7870	1327 i guanidine nucleotide exchange
pir2:F71613	98.00	135.25	0.8422	1121 i hypothetical protein PFB0495w
pir2:A05037	97.00	127.41	1.21	2136 i hypothetical protein 2136 - 11
pir2:T09080	97.00	124.75	1.29	2819 i probable chloroquine resistanc
pir2:S32559	96.50	143.41	0.8984	369 i biglycan precursor - bovine
pir2:S32344	96.50	140.27	0.9685	512 i hypothetical protein 7.9 - Salm
pir2:C71608	96.50	130.64	1.10	899 i origin recognition complx subun1
pir2:H71606	96.00	129.75	1.22	1398 i hypothetical protein PFB0755w
pir2:B42239	96.00	129.75	1.36	1407 i adenylate cyclase (EC 4.6.1.1)
pir2:S520811	95.50	141.75	1.11	369 i proteoglycan I - mouse
pir2:S32793	95.50	141.75	1.11	369 i biglycan precursor - rat
pir2:I51703	95.50	132.64	1.38	954 i c-kit-related kinase 1 (Krk1)

pir2:S37621	+	95.50	130.22	1.46	1228 i Ca2+-transporting ATPase (E
pir2:T09079	+	95.50	122.64	1.76	1182 i probable chloroquine resist
pir2:G71607	+	95.00	129.76	1.62	1182 i probable integral membrane
pir1:BGHUN	+	94.50	140.11	1.37	368 i biglycan precursor - human

seq_name: pir2:A49772

seq_documentation_block:

cuticle protein cut-1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 07-Apr-1994
C:Accession: A49772
C:Sebastian, M.; Jassandro, F.; Bazzicalupo, P.
Dev. Biol. 146, 519-530, 1991
A:Title: cut-1 a Caenorhabditis elegans gene coding for a dauer-specific noncollagene
A:Reference number: A49772; MUID:91323673
A:Accession: A49772
A:Status: preliminary: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-423 <SEB>
A:Cross-references: GB:M55997

alignment_scores:

Quality	Ratio	Length
1379.00	4.191	415
Percent Similarity: 79.277		Gaps: 6
Percent Identity: 65.783		

alignment_block:

US-09-323-427-1 x A49772 ..

Align seg 1/1 to: A49772 from: 1 to: 423

191	TCCTACACCTATTCGATTCGCTTATTCG...	ATTCGCTTACACATCG	237
8	CysLeuAlaAlaLeuValLeuSerAlaSerAlaIleProValAspAsnS		24
238	TCCTGAGGCTGACGAGAAATGATGTCGACCAATGTCGACCAATCA		287
24	ValGluProGluGluGluValGluGluGluGluGluGluGluGluGlu		41
288	ATTTTAATACAGTAATGATTCGACGACATGTTATGTAAGGCTTT		337
41	snRhehnRhrgrsnpRhehnRhrgrsnpRhehnRhrgrsnpRhehnR		57
338	TATGATCAAGAGGTTGCCGATGATGATGATGATGATGATGATGATG		387
58	TyrAspGlnAlaGlyCysArgSerAspGluGluGluGluGluGluGlu		74
388	AATTTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC		437
74	YlleguLeuProPheAspSerCysAsnRhrAlaRhrAlaRhrAlaRhr		91
438	ATTCACGTCGATTTTTCATACACACACTGTTGTCATTCGTTATCA		487
91	snRProGlySerValPheValSerThrValValIleSerPheHisPro		107
488	TATATGTACCAAGTGTGATGATGATGATGATGATGATGATGATGAT		537
108	GlnPheValThrLysValAspArgAlaTyrArgIleGlnCysPheTyr		124
538	GGAAGCTGATAAAGAGTATGTCAGATGATGATGATGATGATGATGAT		587
124	tGluSerAspLysThrValSerThrClnIleGluValSerAspLeuThr		141
588	CTGCTTTTAAACTCAATGTCGATGTCGATGTCGATGTCGATGTCGA		637
141	hValAlaPheInThrGlnValAlaPrometProValCysLysTyrGlu		157
638	TTCGATGTCGACCAACCGTCACACAGTTCATTCATTCATTCATTC		687
158	LeuAspGlyGlyProSerGlyGlnProIleGlnPheAlaThrIleGly		174

```

688 GCCAGTTTATCATTAATAATGAGCATGGAGTTCTGGAACCGTTATATCTTCT 737
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
174 nclnAlvalYrHlslYsrPThrCysAspSerGclurHrThraspHrPhec 191
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
738 GCGCGGTTTCATTCCGCTTTCGTCGATGATGGAACCGATGATCTGTG 787
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
191 yAlAlvalYAlHlsserGysThrValAlapaspGlyAsnGlyAspThrVal 207
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
788 GAAATTCCTAAATGCTGATGATGCTGCTCTTGATTAATATTTGGCTAAAT 837
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
208 GlnlleleuAsnGlnlGclucYsAlAlAlAlAlAlAlAlAlAlAlAlAlAlAl 224
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
838 TTTTGGAAATTCACACAGATTTAATGGCTGGCCAGAGACCTACGATATACA 887
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
224 nleuGlnYrProrHrAspLeuMetAlaGlyGlnGlnAlaHlssValYrL 241
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
888 AATPATGGGATGATGACACAGCTTTTCTATCAATGCGACATGATGATTACC 937
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 ysrYrAlAlAspArgSerGlnleuPheYrGlnGlyAsnGlnlleSerlleThr 257
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
938 ATTAAAGAACCAAAATAGGAAATGTGTGTGACACACAAATGTTTCAGAACACA 987
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
258 lleYsAspProGlnYserGlnGlyAsnAlArpProHrThrcysSerGlnProGln 274
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
988 AGCATTCGAGCTGTTAAACACAGGTGGTGGCCGACGACAAACCTCTGCAG 1037
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
274 nGlnPheGlnYAlAlAlAlYsGlnAlAlGlyAlGlyAlAlAlAlAlAlAl 291
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1038 CT..... 1039
      ||| ||
291 lAlAlArpGlnAlAlGlyAlGlnGlnAlAlAlAlAlAlAlAlAlAlAlAlAlAlAl 307
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1039 ..... 1039
308 AlArpOvalAlAlAlArpOvalAlAlAlAlAlAlAlAlAlAlAlAlAlAlAl 324
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1040 ..... GCGCAACTCTGTTACTC... AAGAAAGATCT... G 1070
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
324 oArGAlArHrleuAlAlGlnleuArGlnleuAlArGlyslYsArgSerPhec 341
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1071 CAGAACCGGAGAAATTCATGATGATGACGACGATTCACACACCTGTGA 1120
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
341 lYGlAluAsnGlnYlleleuAspValAlArGValAlGlnlleAsrThrleuAsp 357
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1121 ATTAGGAGATATACCAAGCTTGCCACTGATTAATTACGTCACCGTGACT 1170
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
358 lleMetGlnGly..... AlAserProSerAlArpGlnAlAlAlAlAlAlAl 372
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1171 TCTGCACATTAATGACACACCTGTATATCTGCTGCACGACACAAATGGA 1220
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
372 lValleSerGln..... GlnSerValAlArGAlArGAlAlArHrSerThrGlyI 387
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1221 TCTGCATGTCACCAATTGGCTTCATCAAGTTTATGAGGTTTAAGCATTCGA 1270
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
387 leSerSerThrProleGlnYleuProSerPheleuGlyMetArGThrIle 403
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1271 TTGATTTGCTGCGCATTAATTAACCTTTCTGTTTAATTTCTGCA 1315
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
404 ValAlAlArHrAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAl 418
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
name: plr2:S27799

```

```
seq_name: p1R2:S27799
seq_documentation_block:
cuticlin 1 - Caenorhabditis elegans (fragment)
C:Species: Caenorhabditis elegans
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Sep-1997
C:Accession: S27799
R:Sebastiano, M.; Lassarro, F.; Bazzicalupo, P.
submitted to the EMBL Data Library, July 1991
A:Description: CWR-1 a Caenorhabditis elegans gene coding for a dauer specific non colla
A:Reference number: S27799
A:Accession: S27799
A:Molecule type: DNA
```

A:Residues: 1-308 <SEB>
A:Cross-references: EMBL:M55997, NID:g156271, PID:g156272
C:genetics:
A:gene: CUT-1
A:introns: 245/3

alignment_scores:	
Quality:	897.50
Ratio:	3.989
Percent Similarity:	74.503
	Percent Identity:
	60.536

```
alignment_block:
US-09-323-427-1 x S27799 ...
```

Align seg 1/1 to: S27799 from: 1 to: 308

527 TGCTTTACATGAGACGATGAAAAAGTATGACACAGATTTGAGTATTC 576
|||||
6 CysPheTyrMetGlnSerAspLysThrValSerThrGlnIleGluValSer 22
|||||

577 TGAATCACCAACTGCTTTTCAAACTCAAATTGTCCCGATGCCAGTATGCC 626
|:::|||||||:::|||||||:

627 GTTATGAATTTTGGATGGTGGACCAACCGGTCACACAGTTCATTGCT 676

39 ySTyrgIuIleLeuAspGlyGlyProSerGlyGlnProIleGlnPheAla 55

677 ATCATTTGGTCAGCCAGTTTATCATTAATGGACATGGGATTCTGAACCGT 726

56 ThrIleGInGlnValTyrHisLysTrpHisCysAspSerGluThrTrh

727 TGATACTTCTGCGGGTGTCCATTCCTGCTTGTGCGATGATGGTAACG 776

72 rasptThrphcysalavalHisSerCysThrvalaspaspGlyasnG 89

777 GTGATACTGTGAATTCCTAAATGCTGATGATGCTGCTTGATAAATAT 826

89 LyspThrValGlnIleLeuAsnGluGluGlyCysAlaLeuAspLysPhe 105

827 TTGCTAATTAATTGGATATCCACACAGATTTAATGGCTGGCCACAAGC 876

106 LeuLeuAsnAsnLeuGluTyrProThrAspLeuMetAlaGlyGlnGluAla 122

877 TCACGTATACAATATGCGATCGATCAGCTTTCTATCAATGCCAGA 926

122 ahisvaltyrlstyalaspargserglnleuphetyrclncysclnt 139

927 TCAGTATTACCATTAAGAACAATAGCGAATGTGTTCCGCCACAATGT 976

139 IeserIleThrIleLysAspProGlySerGluCysAlaArgProThrCys 155

977 TCAGAACCAAGGATTCGGAGCTGTTAAACAGGTGTCGCCGACGCAA 102

156 SerGIuPROGIuGIypheCIyAlaValLysGIuAlaGIyAlaGIyGIyAl 172

1027 ACCTGCTGCAGCT..... 103

172 AHISAlaAlaAlaProGlnAlaGlyValGluGluValGlnAlaAlaP 189

1039	103
------	-------	-----

189 rovalGlyalalProvalAlalalProvalAlalalalalalAla 205

1040GGGCACTTCGTTACTC..AAGAA 106

206 ProAlaValProArgAlaThrLeuAlaGlnLeuArgLeuLeuArgLysLys 222

1063 AAGATCT..GCAGAACCGAGATATCATTTGATGTACGACTGATATCA 110

222 SARGSERPHEGLGLUASNGLGLYLELEUASPVALARGVALGLULEA 239


```

363 ephelysleuileleasnphelyslystyrileasnillelysphesnsl 380
1171 TCTGCACATATGACACACCTGTAATAC.....TTGCTG 1205
      ::::::::::::::: |||
380 eullepheleuMetasnllylleuetyrsnlyrsnasnilleleuphe 396
1206 CAGTACAAATGGAATCGCATGTCACCATTTGGCTTCGAATGTTATG 1255
      ::::::::::::::: |||
397 Glutyllys.....Tyrilleuglnasnlnlntyrillely 408
1256 GGTTAAGCATTTGATTTGCTGCGCATATTATACATTCCTTCGTTAA 1305
408 scysasnphelletyrsnserlleserlysnphelyslyrsnleua 1425
1306 ATTTGCTGCAATCAGAGGCATAAAATATGTTAGAAATCATGAGCA 1355
      || :::: || ::::: ||| ||| |||
425 snasnile.....llellystyrleuasnlnval..... 435
1356 ATATATAACTGCCATATATTCGTTTCTTTCATCCTCTTAATA 1405
      :: ||| ||| ::::: ||| ||| ||| |||
436 .....llelystyr.tyrasnlyserasnilleglnleulel 449
1406 CTATATTTACCTACAAATATATAGTAGAGAAATATTAAGTAATA 1455
      ||| :::: ||| ||| ::::: ||| :::: |||
449 ysasn.llellysnlystrilleuetyrsn.....lletyrthly 463
1456 CAATAAGTGATATTTTCATCAAACTCTTCATCGCTT..... 1495
      | ||| :::: ||| ::::: ||| ::::: |||
463 ltyrleutytyrtyrhisllelyspheetyrsnleutyrsnlysgly 480
1496 .....TATAGCTTCTGA 1507
      ||| |||
480 llelleuasnasnasnasnlystyrasnvallelyrphleulle 496
1508 AAGCTTATTCATTTACATCTTTATATGCA..... 1543
      :: :::: ||| ||| ||| ||| |||
497 Asnlyrphesnleupheserasnlytyrtyrlyslleetyrsnasnas 513
1544 .....TACTATTTGTAAA.....TGTTTCATCA 1565
      ||| ||| ||| ||| ||| ||| |||
513 ntyrasnphelleasnserasnlytyrtyrphelysllysmetasnphel 530
1566 TTAGGCCATGAATAGTTTCGTTGTTATTCATCATTAATCAACTGTGC 1615
      :::: ||| ||| ||| ||| ||| ||| |||
530 enulys.....Asnphasnasnilleglnlleuasnlyslleu 542
1616 TATTTATTTCTACAGTTTATCATTTGTGATATATCACAATTAACCT 1665
      ::::::::::::::: ||| ||| ||| ||| |||
543 PheTyValasnasnillephele.....Tyrtyrlys.TyrGlnL 556
1666 TGATATGCCCAATTTTATGCGCATCATTTCTATTCGTAACAATTC 1715
      ||| ::::: ||| ||| ||| ||| ||| |||
556 ysllyseupheilletyrleuasnillele.....Asnasn... 567
1716 CTATTTGTCATTTGCAATTAAGAATAT 1745
      ||| ||| ||| ||| ||| ||| |||
568 .....llellellysllystyr 573
seq_name: p1r2:F71606
seq_documentation_block:
Probable multiple transmembrane domain protein PFB0770c - malaria parasite (Plasmodium f
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 07-May-1999
C:Accession: F71606
R:Gardner, M.J., Tettelin, H., Carucci, D.J., Cummings, L.M., Aravind, L., Koonin, E.V.,
: Pertea, M., Salzberg, S., Zhou, L., Sutton, G.G., Clayton, R., White, O., Smith, H.O.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743
A:Accession: F71606
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

```

```

A:Residues: 1-1122 <GAR>
A:Cross-references: GB:AEO01417; GB:AEO01362; NID:g3845271; PID:g3845273; TIGR:PFB077
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0770c

alignment_scores:
      Quality: 116.50      Length: 508
      Ratio: 0.492
Percent Similarity: 46.654      Percent Identity: 21.654

alignment_block:
US-09-323-427-1 x F71606 ..

Align seg 1/1 to: F71606 from: 1 to: 1122

376 TCAAGTTCGGCGAATTCACCTTCATTCATTCATCAATGTCGCGTA 425
504 AsnThrCysasnlnhlythr.....Cysasnlnh 514
426 CACGATCTCGAATCCACGTCGTATTTGTAAACAACATGTTGTCAT 475
514 strThrCysasnlnhlythr.....CysasnlnhlythrCysasn. 527
476 TCGTTTCATTCATTTATTTGTTACCAAGTTGATCGTCATATCGAGTACA 525
528 .....AsnhlythrCysasnlnhlythr 535
526 ATGCTTTTACATGAGCTGATTAACAGTATGACAGATTTAGAGTAT 575
536 CysasnlnhlythrSer...AspnasnlnhlythrCysasn..... 547
576 CTGAAATCACAACCTGTTTCAAACTCAAAATGTCGCCGATGCCATATGC 625
548 ...AsnhlythrCys...AspnasnlnhlythrCys..... 556
626 CGTTATGAATTTTGGATGCTGACCAACCGGTCAACAGTTCAATTCG 675
557 .....AsnlnhlythrleuGlyasnpro 564
676 TATCATTTGGCAGCCAGTTTATCATTAATGACATGCGAT..... 715
565 Hisphetyrsn.ProhlysnphetyrsnlnhlyleuasnmetProasnA 581
716 ....TCGAACCGTTGATACCTTCTGCGCGTTGTCATTCCTGCTTT 760
581 snlyslsgluthrhisasnphle..... 589
761 GTGATGATGCTGACCGGTGATCTGGAA.....ATTCTAATGC 801
590 .....SerhisasnlnhlythrGlnGlyasnlnleuetyrsnly 603
802 TGATGATGTCGCTTGATTAATATTTGCTAATAATATTTGGAATATCCA 851
603 snspgLy.....leutyrsnlnhlysllyserlytyrAspA 615
852 CAGATTTAATG...GCTGGCAAGAAGCTCACGTAATACAAATATGCGAT 898
615 snasnleupheGlyala.SerasnlyslleuthSerhisGlnasnll 631
631 eulyslyllelleGlnleuasnlnhlythrlyslleuValGlnGlyargAsnA 648
949 AATATACGATGTGTCGACCAAGATGTCACAGACACACAGATTCGGAG 998
648 sn.....serleuasnlnhlythrlysllytyrAsnlnhlyserasnA 662
999 CTGTTAAACAGAGTGTGCCGACGACAAACCTGCTCAGCTCGCAACTT 1048
663 leuasnlnlytyrPhe.....AspnasnlnleuGlnasnlnlele 676

```

```

1049 CGTTTACTCAAGAAAAGATCTGCAGACCGGAGATA..... 1085
      : ||| : : : : : : : : : : : : : : : : :
676 uSerTyrIleuGlyMetAsnIleLeuYAsnLysIleGlyThrLysPheI 693
1086 .....TCATTCATGTACACACTGATATCAACA..... 1112
      : : : : : : : : : : : : : : : : : :
693 IeMetAsnLysLeuMetYThrAsnValSerAsnAsnGluArgTyrArg 709
1113 .....CCCTGAAATTCAGATGATATCAACACTTCGCCA 1147
      : ||| : : : : : : : : : : : : : : : :
710 TyrTyrIleuAspAsnLeuLysValSerYrIleAsnGlnLeuArgLe 726
1148 GTTGATT...TACGTACCCGTGCACCTCTGCACATA...ATGACAACC 1191
      : : : : : : : : : : : : : : : : : :
726 uMetIleProYrIleThrYrCysLeuGlyLysIleAlaMetSerIleV 743
1192 TGTATACTCTGCGATGACAAATGGAATCGCATGTCACCATTTGGCT 1241
      : : : : : : : : : : : : : : : : : :
743 aLpheTyrIlePheYrIleLysPheAspIleSerYrIleuLysLeuIle 759
1242 TCTCAATGTTATGCGTTTAAGCATTCGATTCGTGCCCTCATTTAT 1291
      : ||| : : : : : : : : : : : : : : : :
760 .....LeuTh 761
1292 ACCATTTGCGTTTAAATTCGTCCAATGAGAGCATAAATAATGTTA 1341
      : : : : : : : : : : : : : : : : : :
761 rAspYrLysMetYrPhe...LysLeuPheGluHisLysAsnIleIleP 777
1342 GAATCATGAGCAATATATTAACCTGCCA.....TAT 1373
      : : : : : : : : : : : : : : : : : :
777 heIleValSerPheSerIleIleLeuGlyAsnThrIleIleSerPheP 793
1374 ATATTGTTTCTTATCATCCTCTTATATACATATTTCAGTAACAA 1423
      : : : : : : : : : : : : : : : : : :
794 SerPheIle..PheLeuSerSerPhePheGluValIleuSerThrIleu 809
1424 TATATATGATGTAGG.....AATAATTAAGTGT 1452
      : : : : : : : : : : : : : : : : : :
810 PheIlePheIleLysCysIleSerGluPheLeuPheLeuLeuValVa 826
1453 ATACATAAGTCATAT.....TTTCATCAAAAC..... 1480
      : ||| : : : : : : : : : : : : : : : :
826 ITrAsnGluValPheGluIlePheLeuArgAsnIleLysGlnProAspL 843
1481 .....TTCTTATCGCTTTTATAGCTCTGAAAAGCTATTATCA 1519
      : : : : : : : : : : : : : : : : : :
843 ySTyrAlaProYrPhePheLeuThrPheAlaValIleProSerPheLys 859
1520 TTATTCAGTATCTTTTATATGATACATATTTGTAA..... 1555
      : : : : : : : : : : : : : : : : : :
860 IleIleArgAsn...IleYrPhePheLysCysAlaLeuSerGlyArgG1 875
1556 .....TGTTCATCATTTAGGCCATGAATAGTTTCGTTTATTATATCA 1598
      : : : : : : : : : : : : : : : : : :
875 nPheIleLeuYrIleIleArgProPheIleLysAsp..... 887
1599 TCATATCAACTTGCTATTATTATTAACAGTTTATCATTTGTATTA 1648
      : : : : : : : : : : : : : : : : : :
887 ..... 887
1649 TATCACAATATATACCTGTATTTGCCAATTTTATGGGATCATTTCT 1698
      : ||| : : : : : : : : : : : : : : : :
888 .....LysAsnIleSerLysLeuProAsnPhePhe..AsnIleLysGluT 902
1699 ATTCGTGTAACAATTCA 1715
      : : : : : : : : : : : : : : : : : :
902 yTrAsnAsnAsnAsnAsn 907
seq_name: p1r2:B26696
seq_documentation_block:
hypothetical protein 1 (cyb-COII intergenic region) - Leishmania tarentolae mitochondria
C:Species: mitochondrion Leishmania tarentolae

```

```

C>Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 24-Sep-1999
C:Accession: B26696
R:Simpson, L.; Neckelmann, N.; de la Cruz, V.F.; Simpson, A.M.; Feagin, J.E.; Jaszmer,
J. Biol. Chem. 262, 6182-6196, 1987
A:Title: Comparison of the maxicircle (mitochondrial) genomes of Leishmania tarentola
A:Reference number: A92643; MUID:87194837
A:Accession: B26696
A:Molecule type: DNA
A:Residues: 1-443 <SIM>
A:Cross-references: GB:M10126; GB:J02707; GB:M10127; GB:M1022; GB:M64690; GB:N00030;
A:Note: the authors translated the codon ATT for residue 388 as Phe and TTC for resid
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC6
C:Superfamily: hypothetical protein 1 (cyb-COII intergenic region)
C:Keywords: mitochondrion

alignment_scores:
      Quality: 112.00      Length: 544
      Ratio: 0.463      Gaps: 29
Percent Similarity: 44.485      Percent Identity: 19.853

alignment_block:
US-09-323-427-1 x B26696 ..
Align seg 1/1 to: B26696 from: 1 to: 443

180 TTATTGCTTTCTGTACTACACTTATTCATTTGCTTATTCGATTCGCGT 229
   |||:||||| : : : : : : : : : : : : : : : :
3 LeuPheLeuYrIleuIleHisIleIleLeuPheLeuLeuYrSerP 19
230 GACAAATGCGTGAAGGTGAGCCAG.....AATTTGAATGTGAGCAAC 273
   : : : : : : : : : : : : : : : : : :
274 TTCAATTAACAATCAATTTTATATACAGTAATGATTCAGAGCATGTTT 323
   : : : ||| : : : : : : : : : : : : : : :
36 IeTrpLeuIleIleLeuAsnIleIleIleLeuThrIleLeuAspSerYr 52
324 ATGTGAAGGCTTTATGATCAAGAAAGTGGCGGTATGATGATGAAGTGA 373
   : : : : : : : : : : : : : : : : : :
53 IleCysPheIlePheLeuLeu..... 59
374 CGTCAAGTGTGCGGAATTTTCACCTTCATTTGATTCATGATGATGGCG 423
   ||| : : : ||| : : : : : : : : : : : :
60 .....LeuPheLeuPheCysPhePheLeuPheCysPheLeuAsnP 74
424 TACAGATCTCTGAATCCAGCTGATTTTGTAAACAACAACCTGTTCGA 473
   : : : : : : : : : : : : : : : : : :
74 heAspThrArgPheValPheMetIleIleIleMetGlnTrIleIle 90
474 TTTCGTTTCATCATTTATTTGTACCAAGTTGATCGCATATGAGTA 523
   ||| : : : : : : : : : : : : : : : :
91 PheMetPheLeuHis.....ValIleHisIleLeuPhe 101
524 CAATGCTTTTACATGAGACGTGATAAACAAGTTAGGCACAGATTGAGGT 573
   : : : : : : : : : : : : : : : : : :
101 eIleSerIleLeuPheGlnLeuPhe..... 109
574 ATCTGAATACACAACCTGTTTCAACATCAATATG...TCCGATGCCAG 620
   ||| : : : ||| : : : : : : : : : : : :
110 .....SerLeuLeuPheLeuIleLeuIleSerSerArgPheGly 123
621 TATGCCGTTATGAATTTTGGATGGTGACACACCGGTCAACCATTCGA 670
   ||| : : : : : : : : : : : : : : : :
124 TyrLysIleLeuIleLeuThrPyr.....TyrTyrMetIleAs 136
671 TTTCGATATGATGCGACGACGATTTATCATTAATGACATGCGATTCGA 720
   ||| : : : : : : : : : : : : : : : :
136 nLeuIleAsnPheIleLeuLeuPheValLeuLeuTyrTrMetIleLeu 152
721 AACCGTTGATCTTCTGCGCGTGTGCCATTCCTGCTTTGTC..... 763

```



```

135 nasnsnlysasnilletyValtyriletyrileserleuglyLeuA 152
136 ..... ||| ..... |||
137 322 TTATGTGAAGGCTTTATNGATCAGAGGCTGCCGTATATGATGAGGCG 371
138 : : : : : ||| ||| ||| : : : : :
139 152 rgile.....llephelysileserlys.....leuasnile 162
140 : : : : : ||| ||| ||| : : : : :
141 372 GACGTCAAGTGGCCGAATTCCTCACTTCATTTGATTCATGCAATGTTCG 421
142 : : : : : ||| ||| ||| : : : : :
143 163 AspGyTYrPheasnlnphelysPheasnphelleulelleuleuleuty 179
144 : : : : : ||| ||| ||| : : : : :
145 422 CGTACACAGATCTGTGAATTCACACGCGTATTTTGTACACACAACGTGGT 471
146 : : : : : ||| ||| ||| : : : : :
147 179 rleuasnln.....lletyrileasnlnlyslnleule..... 204
148 : : : : : ||| ||| ||| : : : : :
149 472 CATTCGTTTCATCCATATTTGTTACCAAGTGTGTCGATTCGATTCGAG 521
150 : : : : : ||| ||| ||| : : : : :
151 191 eupheilletyrlnasnlnlelleasnlnlyslnleule..... 204
152 : : : : : ||| ||| ||| : : : : :
153 522 TACATGCTTTACATGAGAGCTGATTAACAGTTAGTGACAGATGAG 571
154 : : : : : ||| ||| ||| : : : : :
155 205 TYrAsnTYrile.....lyspheilletyrserlystyrsnlnleas 219
156 : : : : : ||| ||| ||| : : : : :
157 572 GTATGTGAATTCACACGCTTTTTCAAACGCAATGTCCGATGCCAGT 621
158 : : : : : ||| ||| ||| : : : : :
159 219 nasnilleiserleuylsleupheillelylsleuasnlnlysPheasnA 236
160 : : : : : ||| ||| ||| : : : : :
161 622 ATGCCGTTATGAAATTTGGATGGTGAGACCAACCGTCAACCACTTCAT 671
162 : : : : : ||| ||| ||| : : : : :
163 236 sniletyr.....lleasnlnleuasn 243
164 : : : : : ||| ||| ||| : : : : :
165 672 TTGCTA.....TCATGTGCAGCCAGTTATTCAT 700
166 : : : : : ||| ||| ||| : : : : :
167 244 lleuphieserillelysleuasnpheserlytyrserAspPheTYrll 260
168 : : : : : ||| ||| ||| : : : : :
169 701 AAATGGA.....CATGCAATCTGAAACCGTTGAPAC 732
170 : : : : : ||| ||| ||| : : : : :
171 260 easnlnletyrlnasnlnlysPheTYrserilleleasnlnleuleu 277
172 : : : : : ||| ||| ||| : : : : :
173 733 TTTGCGCGGGTTCACATTCCTGCTTGTGCATGATGATAGGTAGATA 782
174 : : : : : ||| ||| ||| : : : : :
175 277 lelyserlys..... 280
176 : : : : : ||| ||| ||| : : : : :
177 783 CTGTGGAATTTCTAATGCTGATGATGCTCTTGATTAATATTTGCTA 832
178 : : : : : ||| ||| ||| : : : : :
179 281 .....LysTYrleuylsillephelystyrglnleuleAsnilleasnAr 295
180 : : : : : ||| ||| ||| : : : : :
181 833 AATATTTGGAATATCCACAGATTTAATGGCTGCCAAGAAGCTCACGT 882
182 : : : : : ||| ||| ||| : : : : :
183 295 gasnilletyrln.....Asnillethyleuleuleuasnlnlyst 310
184 : : : : : ||| ||| ||| : : : : :
185 883 ATACAAATATGGGATGATGCATCACACTTTCTATCAATGCCGATCAGTA 932
186 : : : : : ||| ||| ||| : : : : :
187 310 Yrllleasnillelleuclu.....AsnilleasnilleasnProleu 323
188 : : : : : ||| ||| ||| : : : : :
189 933 TTACCAATTAAGAACCAATATGCAATGTGTTCGACCAAT..... 974
190 : : : : : ||| ||| ||| : : : : :
191 324 ValGlnTYrserAspGlnValasnlnleuserGlnlleasnGlnlyPh 340
192 : : : : : ||| ||| ||| : : : : :
193 975 .....GTTCAGAACC 984
194 : : : : : ||| ||| ||| : : : : :
195 340 elysilleasnmetilethrhglyleuasnserlyspheilleuAsnA 357
196 : : : : : ||| ||| ||| : : : : :
197 985 ACAAGAGTGGAGCTGTTAAACAGGTGGCCGACAGAAACCTGCTG 1034
198 : : : : : ||| ||| ||| : : : : :
199 357 snAspleuArgGlnleuProArgasnilleuGlyTYrilleserleu 373
200 : : : : : ||| ||| ||| : : : : :
201 1035 CAG.....CTGCCCAACTTCGTT..... 1052
202 : : : : : ||| ||| ||| : : : : :
203 374 AsnThrAsnGlnGlyleuThrcysGlyleuValAsnTYrleuThrhAs 390
204 : : : : : ||| ||| ||| : : : : :
205 1052 ..... 1052

```

```

390 nilepheleuasnleuylstyrlsleupheValleTYrTYrlyshsilleP 407
1053 .....TACTCAAGAAAGATCTGCAGACGGAGCAATATCATTTG 1091
407 heTYrAsnArgTYrAsnphelysleuleuleuAsnillephesnlnysn 423
1092 ATGTACGAACCTGATATCAACACACCTTGAATTAAGCATGATATCAACGT 1141
424 PheTYrAsnilleserPheasnlnleTYrleuLysAsnilleasnPh 440
1142 TTGCCAGTTGATTTACGTCACCGTCAC.....TTTCGCA 1176
440 easnlysthThrhilleleuThrhleasnlnlysasnThrhPheylsillecysA 457
1177 ACATATGAGACACCGTAAATCTTGCTGCAGCTACAAAATGCAATCTGCA 1226
457 snllethGlnasnilleleTYrile.....Pro 466
1227 TGTCACCAATTTGCTTCCAATGTTATGGCTTAAAGCATTCGATTCGATT 1276
467 PheAsnTYrleuLeuser.....Ph 473
1277 GCTGCCGTATTTATTCACATTTGTTAAATTTGTCACAAATCAGAGGC 1326
473 elleglnasnleuileProPhe...llehistryasnspserilleArgA 489
1327 ATAAATAATATTTAGATCATCGAAGCAATATAAACTGCCATATATA 1376
489 snLeuMeiserillelysmethlnsthrglnllevalProilleleTYrPro 505
1377 TTGCTTTCTTCATCAATCCCTTAATACATTAATTTGCTCAAAATAT 1426
506 AsnLeuserAsnile.....llethAsnly 514
1427 ATAGATGTAGAAATTAATCTGTAATACATA.....A 1461
514 rAsnphelleleuasnlnlystyrlsleuasnlnleulelleserTYrgln 531
1462 GTGATATTTTCACAAACTCTCTATCGCTTTATAGCTTCGCAAG 1511
531 lUGlylleValleTYrValserCysllelysllelleleArgasp... 546
1512 CTATTCATTTATTCAGTAACTTTATATGACTACTAT..... 1550
547 leupheasnArgGlnilleleTYrlyleuasnlnlystlyslleas 563
1551 .....G 1551
563 nGlnasnilleuleuleuleTYrlylspProillevalrpyalglylulysv 580
1552 TAAATGTTTCATCATTTAGGCCATGAAATAGTTTCGTTGTTATTCATCA 1601
580 alAsnillelyGlnilleleuAlalleasnSerAsnleuasnserGln 596
1602 TTATCAACTTGTCTATTTATTTATTAACAGTT..... 1634
597 TYrserleuGlyAsnasnleuuleuValglyTYrlyserlyleuGlyTY 613
1635 .....ATCATTGTGATATATATCAACAATTTACTCT 1665
613 rGlnTYrGlnAspAlallellellelelelelelelelelelelelelele 630
1666 TGTATTTGCCCAATTTTATGGGCACTTTTATGCTATTCGTAAC..... 1709
630 euryThrhserleuhsleuasnlnleTYrGlnlleserleuasnillele 646
1710 AATTCACCT.....ATTGCAATTAATGCAATTAATAAGTATTTTC 1748
647 AsnAsnilleProGlnillecysserilleasnleuserlysmetyr 661
seq_name: ph2:S78177
seq_documentation_block:

```

hypothetical protein 717 - RecJinomonas americana (ATCC 50394) mitochondrion
 C:Species: Mitochondrion RecJinomonas americana
 A:Variety: ATCC 50394
 C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 17-Mar-1999
 C:Accession: S78177
 R:Lang, B.F.; Burger, G.; O'Kelly, C.J.; Cedergren, R.; Golding, G.B.; Lemieux, C.; Sank
 Nature 387, 493-497, 1997
 A:Title: An ancestral mitochondrial DNA resembling a eubacterial genome in miniature.
 A:Reference number: S78127; MUID:97311393
 A:Accession: S78177
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-717 <LAN>
 A:Cross-references: EMBL:AF007261; NID:g2258325; PID:g2258376
 A:Experimental source: ATCC 50394
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1997
 C:Genetics:
 A:Genome: mitochondrion
 C:Keywords: mitochondrion

alignment_scores:
 Quality: 108.00 Length: 563
 Ratio: 0.414 Gaps: 28
 Percent Similarity: 46.359 Percent Identity: 19.538

alignment_block:

US-09-323-427-1 x S78177 ..

Align seg 1/1 to: S78177 from: 1 to: 717

```

309 TCGAAGCAGATGTTATGTGAAAGTCTTATGATCAAGAAGGTGCCGT 358
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
126 ThrLysAspLeuIlePheLeuMetPheLeuAsnSprLleLysTyr 142
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
359 AATGATGAGGTGAGGTGAGTGCAGTGGCGAATTCTCACTCCATTGATTC 408
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
142 rAsnLysAsnLeuIleAsnAsnGlnAsnGlnLeuTyrPheHis... 156
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
409 ATGCATGTTGCCGCTACACAGATCTCTGAATCCAGCTGATTTTGTAA 458
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
157 .....TyrPheLeuPheLeuTyr 162
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
459 CAACAACGTGTTGTCATTCGTTTCATTCATTTATTTGTACCAAGTGTAT 508
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
163 AsnAsnIleCysAsnThrAsnTyrVal.....LeuIle 173
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
509 CGTGCAATCGAGTACATGCTTTTACATGGAAGCTGATATAACAGTTAG 558
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
173 eThrLysThrGlnLysAsnSerPheAlaTyr...LeuLeuLysGln...T 188
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
559 TGCACGATTGAGGTATGTAATCACAACCTGTTTCAACCTCAATTTG 608
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
188 hLysAsnLeuGluTyrIleAsnIleGlnThrIleThrLysIleLysGln 204
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
609 TCCCGATGCCAG.....TATGCCGTAT 631
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205 AsnTyrLeuGlnHisLeuPheAsnThrGlnAsnLysTyrGlnLys 221
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
632 GAATTTTGGATGGGACCAACCGTCAACAGTTCAAT..... 671
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
221 nGlnSerPhePheTyrGlnGlnLysPhe.....PheAsnAsnTyrGlnT 236
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
672 .....TTGCTATCATTCGTCAGCCAGTTTATCATTAATGACAT 710
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
236 rPlLysLysGlnIlePhePheThrIleAspGlnIleLeuTyrLysThrHis 252
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
711 GCGATTTGAAACCGTTGATCTTCTGCGCGGTGTCATTCCTGCGCTT 760
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
253 LeuThrAsnLysGlnPhePheProLysGlnIleAsnLeuThrAsn... 268
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
761 GTCGATGATGTAACGCTGATCTGTGAATTTCTAATGCTGATGATG 810
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

269 .....AspValGlnLeuPheLys..... 274
811 TCGCTTGATAATATTTCTAAATATTTGGAATATCCACAGATTAA 860
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
275 .....ThrIleAsnAsnAsnThrThrIleTyrGlnIleLysIleProHis 289
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
861 TGGCTGGCCAGAGAGCTGACGTATCAAAATATGCGATGATCACAGCTT 910
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
290 GlnLeuAsnAsnLysIleLeuTyr...GlnLeuTyrIleGlnAsnAsnI 305
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
911 TTCTATCATTCGACATCAGTATTCATTAAGAACCAATAGCGAATG 960
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
305 nGlnPhe.....LysAsnLysLeuLeuTyrPheLysAsnTyrIleProPhe 320
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
961 TGTGACCCACATGTTTCAGAACCAACAGATTCGAGCTGTTAAACAG 1010
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
320 heLeuGlnLysGlnLysGlnLeuIleAspThrGlnIleLysLys 336
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1011 GTGGTCCCGCAGCAAAACCTGCTGCAG.....CT 1039
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
337 TyrAsnLysLysIleAsnLeuLeuGlnGlnIleTyrTyrProLysVa 353
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1040 GCGCAACTCGTTTACTCAAGAAAGATCTG.....CAGACCGGAGAA 1083
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
353 LLeuAsnTyrGlnTyrPheLysLysAsnLysIleAsnGlnAsnGlnLysI 370
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1084 TATCATGATGATGACGAACTGATATCACACCCCTGAATATACGATGAA 1133
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
370 Le.....IleGlnTyrSerAsnLeuAsnIleGlnLeuAsnThrLeu 383
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1134 ATCAGAGTTTGCAGTTGATTACGTCACC.....GTCCACTTTCGCA 1177
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
384 IleLysTyrIleArgLysThrAsnAspThrPhePheValIleIleLeuAs 400
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1178 CATTAATGACAAACCTGATTAATCTGCTGCAGTACAAAAATGATCGCAT 1227
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
400 nSerLysIleSerLeuLysTyrLeuLeu..... 409
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1228 GTCACATTTGGCTTCTCAATGTTTATGAGTTTAAGCATTCGATTGATTG 1277
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
410 .....AsnIleLeuThrLeu..... 414
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1278 GTGCCGTCAATTATACCATTTGCTTTA.....AATTGCTCCA 1315
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
415 .....GlnAsnLeuSerThrPheLysIlePheLeuGlnAspAsnPheLeuTh 429
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1316 AATPCAGA.....AGGCATAAAATATAGTTAGATCAATCAGTAAGCA. 1355
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
429 rIleGlnThrLeuIleLysAsnLysSerAsnIleIleIleThrAsnAlaG 446
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1355 ..... 1355
446 IncLysLeuAsnArgIleLeuAsnThrAsnThrAsnIleLysAsn 462
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1356 .....ATAATAAACTGCCATATATATTCGTTTCTTATATATCC 1396
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
463 LysIleGlnIleLeuIleLysAspSerLeuPheThrAsnGlnGlnAsnAs 479
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1397 TTCTAATACATAATTTAGCTA..... 1418
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
479 rIleThrAsnLysPheSerLeuTyrIleThrAsnGlnThrLysGlnLysA 496
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1419 .....ACAATATATATAGTATGTAGAAATATATTAATCAATTAAGT 1463
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
496 snIleValAsnIleArgAsnGlnSerGlnLysLysSerAspAsnThrAsn 512
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1464 GATATTTTCATCAAAACTTCTTATTCGCTTTATAGCTTGTGAAGAGCT 1513
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
513 AspPheTyrGlnLeuLeuLysAsnThrGlnLysIleIleAsnLysGlnLys 529
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1514 TATTCATAT...TCAGTAATCTTT.....TATATGCACTACTATG 1551
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
529 rIleTyrTyrLeuAsnIleValPheLeuThrGlnTyrValGlnArgIleA 546
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```



```

1552 TA...AATGTTTCATTAGCCATGATAGTTTCGTTGTATTATCA 1598
      ::::::::::::::: ||||| ||| ::::|
546 laaagaspilleglnlysmetprohlsgluphepilletyrPheunlis 562
1599 TCATTATCACTTGCTCTATTTTATCTACAGTTTATCATTTGTGATTA 1648
      ||:::
563 TyrleuLys..... 565
1649 TATCAAAATATACCTTGATATGCCCCAATTTTATGGCATCATTTCT 1698
      ||| ||::: ||| ||::: |||
566 .....AsnLysTyrleupheCys...iletyrLysGlnAsnThrSert 579
1699 ATTCTGTA.....AACATTCACCTTATTTGTCATTATTCATTA 1739
      ||::: ||::: ||||| ||::: ||::: ||:::
579 yrlvsilleylslsAsnAsnSntYrPheleuTrpGlnSerleuileu 595
1740 AAGTATTTTCATTTGTGAAAAA..... 1778
      :::: ||| ||||| ||| |||
596 Gluilephe.....LysLysLysiletyrLysLys 605
seq_name: pir2:A40970

```

```

seq_documentation_block:
undulin 1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 13-Aug-1999
C:Accession: A40970
R:Just: M.; Herbst, H.; Hummel, M.; Duekorp, H.; Tripieler, D.; Stein, H.; Schnuppan, D.
J. Biol. Chem. 266, 17326-17332, 1991
A:Title: Undulin is a novel member of the fibronectin-tenascin family of extracellular m
A:Reference number: A40970; MOID: 91373351
A:Accession: A40970
A:Molecule type: mRNA
A:Residues: 1-843 <IDS>
A:Cross-references: GB:M64108; NID:g340081; PIDN:AAA36794.1; PID:g340082
C:Superfamily: collagen alpha 1(XIV) chain; fibronectin type III repeat homology: von W
C:Keywords: glycoprotein
F:165-246/Domain: fibronectin type III repeat homology <FN3A>
F:255-338/Domain: fibronectin type III repeat homology <FN3B>
F:347-427/Domain: fibronectin type III repeat homology <FN3C>
F:436-520/Domain: fibronectin type III repeat homology <FN3D>
F:547-632/Domain: fibronectin type III repeat homology <FN3E>
F:641-723/Domain: fibronectin type III repeat homology <FN3F>
F:731-818/Domain: fibronectin type III repeat homology <FN3G>

```

```

alignment_scores:
Quality: 108.00 Length: 396
Ratio: 0.527 Gaps: 21
Percent Similarity: 51.768 Percent Identity: 21.212

```

```

alignment_block:
US-09-323-427-1 x A40970 ..

```

```

Align seg 1/1 to: A40970 from: 1 to: 843

```

```

8 TTACCAAGTTTGAGGTGCTATATACACCGATCGACGACAACAACA 57
  ||| ::::::::::::::: ||| ||:::
228 LeuThrGluTyrGlnIleAlaValPheAlaIleTyrAlaHisThrAlaSe 244
58 CAACAACAACAACAACAACAACAACAATATAATACCCATCAAGTGAG 107
  :::: ||| ||::: ||| :::::
244 rGluglYleuArglYthrGluThrThrLeuAlaLeuPrometaLaseRa 261
108 GAAGAAGACAGAGACATCTTAGTTTCTRAAAATGCAATTTACTAAA 157
  :::: ||| :::::
261 sp.....LeuLeuLeuTyrAspValThrGlu 269
158 TCTTCTGAATGATGATTCGTCTTATGCTTCTTCTGACACCTATATGC 207
  :::: ||| ::::: |||
270 AsnSer...MetArgValLysTrpAspAlaValPProGlyAlaLaseGlyTy 285

```

```

208 ATTGCTTATTTCGATTCGCGTTGACATAGTGTCGAAGT...GAGCAG 254
  ||| ::::: ||::: ||| |||
285 rleuileuTyrAlaProleuThrGluGlyleuAlaGlyaspGluLysG 302
255 AAATGAATGTGGACCAACTTCATATACATCAATTTT..... 292
302 lmetLysilleglYthrHisThrAspIleGluLeuSerGlyleuLeu 318
293 ...ATACACGTAATGCATTCGACAGACATGTTTATGGAAGGCTTTA 339
  ||||| ||| ||| |||
319 ProAsnThr.....GluTyrThrValThrValTyrAlaMetPhe 331
340 TGATCAAGAAGGTGCGTAATGATGAGAAGGTGACGTCAAGTGGCGGAA 389
  :::: ||| ||| :::::
331 eglYgluGluAlaSerAspProValThrGlyGlnGluThrThrLeuAla 348
390 TTTGACCTTCATTTGATTTCATGCAATGTGGCGGTACACAGATCTGTG 439
  :::: ||| ::::: ||::: |||
348 euSerProPro...ArgAsnLeuArgIleSerAsnValGlySerAsnSer 363
440 CCAAGTGTATTTTGTAAACAACACGTGTGTCATTTGCTTCATCCATT 489
  |||
364 AlaArg.....LeuThrTrpAspPro... 370
490 ATTGTTACCAAGTTTGATGTCGTCATATCGATACATGCTTTTACATGG 539
  ::::: ||::: |||
371 ...ThSerArgGlnIleAsnGlyTyrArgIle.....ValTyrAsnA 384
540 AAGCGATAAAAACAGTTAGTGCACAGATTGAGATCTCGAATTCACACT 589
  :::: ||| ::::: ||::: |||
384 snAlaAspGlyThrGluIleAsnGluValGluValAspProleuThrThr 400
590 GCTTTTCAACGCAATATGTCGCGATG.....CCAGTATGCGCGTTA 630
  ::::: ||:::
401 PheProleuLysGlyLeuThrProleuThrGluTyrThrIleAlaIlePhe 417
631 TGAATTTTGCATGCTGACCAACCGGTCAACCACTT.....CAATTTG 674
  :::: ||| ::::: ||:::
417 eSerIleTyrAspGluGlnGlnSer...GluProleuIleGlyValPheT 433
675 CTATATGCTGTCACCACTTATTCATTAATGACATCGGATTCGTAACC 724
  :::: ||| ::::: ||::: |||
433 hrThrGluGluValProAlaGlnGlnTyrLeuGluIleAspIleValThr 449
725 GTTGATCTTTCGCGCGTGTGTCATTCCTTCTTGTCGATGATGTT... 772
  :::: ||| ::::: |||
450 ThrAspSerPheArgValThrThrPheSerProleuSerAlaAspGluLys 466
773 .....AACGCTGATACGTGGAATTC 794
466 whlslsleuMetTrpIleProValTyrGlyGlyLysThrGluGluVal 483
795 TAAATGCTGATGATGCTCTTGATTAATATATTGCTAATATTG... 841
  :::: ||| ::::: ||::: |||
483 alLeuLysGlu.....GluGlnAspSerHisValIleGluLysLeuGlu 497
842 .....GAATATCAACAGATTTAATGGCT.....GGCCA 870
  ||||| ::::: ||::: |||
498 ProGlyThrGluTyrGluValSerleuLeuAlaValLeuAspGlySe 514
871 AGAAGCTCAGCTATACA.....AATATGGAGTCGATCAGCATCTTT 911
  :::: ||| ::::: |||
514 rGluSerGlnValValThrAlaValGlyThrThrLeuAspSerPheThr 530
912 TCTATCATTCGCAGATCAGTATATACATTTAAGAACAATATGCCAATGT 961
  :::: ||| ::::: |||
531 ThrGluProAlaThrThrIleValPro...ThrThrSerValThrSerVa 546
962 GTTCGACCAACATGTTCAAGACCAACAGAGATTCGAGCTGTTAAACAAG 1011
  ||||| ::::: |||
546 lPheGlnThrGlyIleArgAsn.....LeuV 555
1012 TGTGTCGCGCAGCAAAACCTGTGACAGTGGCG...AACTTCGTTACTCA 1058

```

```

||||| .....: .....: |||
555 alvalglyaspplurthrthrserserleuargvallystrpaspilesr 571
1059 AGAAGAATCTGCAGACCGGATATCATGTGATG 1094
572 aspseraspvalglnglnphearvalthrtyrmet 583
seq_name: plr2:C71607

```

seq_documentation_block:

```

hypothetical protein PFB0745w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 07-May-1999
R:Accession: C71607
R:Gardner, M.J.; Tetelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
Pereira, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 287, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743
A:Accession: C71607
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1351 <GAP>
A:Cross-references: GB:AE001415; GB:AE001362; MID:g3845264; PID:g3845266; TIGR:PFB0745w
A:Experimental source: clone 307
C:Genetics:
A:Gene: PFB0745w

```

```

alignment_scores:
  Quality: 107.00      Length: 669
  Ratio: 0.354        Gaps: 38
Percent similarity: 43.832  Percent identity: 19.303

```

alignment_block:
US-09-323-427-1 x C71607 ..

Align seg 1/1 to: C71607 from: 1 to: 1351

```

25 GTCATTACACGACTGCAGCAACAACAACAACAACAACA 74
   |||||: .....: .....: |||
323 lletyrlvsmetasnargllyasnasnlnleghlnhisaspplleas 339
   |||||: .....: .....: |||
75 CAACAACAGCAATAAT.....ACCCCATCAAGTGAAGAGAAGACAG 118
   |||||: .....: .....: |||
339 nasnclurhrasnaspvallyasnasnlnletyrglyarglyslslysa 356
   |||||: .....: .....: |||
119 GAAGCAATCTAGTTTCTTAATAATCGAATTCTTCTTGAAT 168
   |||||: .....: .....: |||
356 snlyasnlnletyrglyasnasnasn..... 365
169 GATGATTCGCTTATTGCTTCTGTACTACACTTATGTCATGCTTATT 218
   |||||: .....: .....: |||
365 .....: .....: .....: 365
219 CGATTCGCGTGTGACATGCTGTGCAAGGTGACCCAGAAATTGATGTGA 268
   |||||: .....: .....: .....: A 366
366 .....: .....: .....: .....:
269 CCAACTTCAATACATCAATTTTATACAGTATGCAATTCGAAAGACA 318
   |||||: .....: .....: .....: .....:
366 snasnasnasnasnasn..AsnlyshsmetasnlysserlleSerThrAs 382
319 TGT.....FATGGAAGGT.....CTTATGATCAAGAAGGT 353
   |||||: .....: .....: .....: .....:
382 ntleuasnlystyrllelysnnglnhisllevalthrlyshlsvali 399
354 GCCGTAATGATGAAGGTGACGTCAAGTTGCCGAATTCACCTTCCATT 403
   |||||: .....: .....: .....: .....:
399 leatgThrAspplurlyslslyslguleupe..... 409
404 GATTCATCAATGTTGCCGCTACAGATCTCTGAATCCAGCTGATTTT 453
   |||||: .....: .....: .....: .....:

```

```

410 ...phecysthrphevalasnmethrthrleu..leutyrgluleilel 425
454 TGTACACAACTGTTGCA.....TTTGGTTTC 482
   || .....: .....: .....: .....:
425 euphetryllysnlnleSerThrAsnlnllelyslleasnlyrolyur 441
483 ATC.....CATTTTGTACCMAAGTTGATGTCGATATGCA 520
   || .....: .....: .....: .....:
442 lleaspserthrtrpAsnlnlelethrAsnlnlellelelyllely 458
521 GTACATGCTTTACATGAGCTGATAAACAGTTAGTCGACACATATGA 570
   || .....: .....: .....: .....:
458 s...AsnlnlePromeGluArgllelyslslyslurhrhs..... 471
571 GGTATCTGAATACACACTGCTTTCAAACTCAATGTCGCCATGCCAG 620
   |||||: .....: .....: .....: .....:
472 ....leughlnserlleleSerleuTySerleuThrvalleuasn 486
621 TATGCCGTTATGAAA.....TTTGGATGTGACCAACCGGTCA 661
   |||||: .....: .....: .....: .....:
487 TyrSerlyslleuTygluAsnlnlephetyrillephegluArgserValas 503
662 CCAGTTCAATTTGCTATCATGTCGACCCAGTTTTCATAATG....A 707
   || .....: .....: .....: .....:
503 pllellehisaspheupheylshsasnmetarglyslleasnlnleket 520
708 CATGCCATTTGCAACCGTTGATCTTCTGCGCGGTGTCGATCTGCG 757
   || .....: .....: .....: .....:
520 hrpheasppluleuLyasnspheuasnvalSerphevalasnmetcys 536
758 .....: .....: .....: .....:
537 AsnaspaspasnasnasnasnasnspaspaspasnnglyAspaspas 553
787 GGAATTTCTAATGCTGATGATGCTCTTGATTAATTTGCTAATA 836
   || .....: .....: .....: .....:
553 pasnasnasnasnaspaspasnnglyAspasp.....Asna 566
837 ATTGGAATATCCACAGATTATAGCTGCGCAAGAGCTGACGTATAC 886
   || .....: .....: .....: .....:
566 sn.....Aspaspasnasnvalille 572
887 AAATATGCGGATGATCAACGCTTTCTATCAATGCCAGATCATGATTAC 936
   |||||: .....: .....: .....: .....:
573 lystyr..lys.....HisSer...AsnvalgluprolustyrAsnL 586
937 CATTAAGAACCAATAGCAATGCTGCGACCAACATGTTGACAGAACAC 986
   |||||: .....: .....: .....: .....:
586 ysvallystyrAsnmettyrAsntrhrphehisargasnlnlelyshelys 602
987 AAGATTCGAGAGCTGTTAAACAGAGGTGGCGGCAAAACCTGCTCA 1036
   |||||: .....: .....: .....: .....:
603 tyrllyslnglnasnlnlevalhisasnlyrleuasnlyslleasproleule 619
1037 GCTGCGCAACTTC.....CTTACTCAAGAA 1062
   |||||: .....: .....: .....: .....:
619 utyrasnasnphleupephevaltyrvalproaspheuleutyrserGlna 636
1063 AAGATCTGCAGACCGGAGCAATATCATGATGTAAGCACTGATATACCA 1112
   |||||: .....: .....: .....: .....:
636 spasnCysThrAsp..... 640
1113 CCCTTGAATATTAGCATGATATCAACGTTTGGCAGTTGATTAAGTCAC 1162
   |||||: .....: .....: .....: .....:
641 .....MetphethrleuaspluleuthrlyslleuTyralale 654
1163 GGTGACTTCTGCACATATATGACACACCTGATATTCTGCTGCATACA 1212
   |||||: .....: .....: .....: .....:
654 userlytyr.....Glnlysg 660
1213 AAATGATCT.....GCATGACACATTTGGCT 1241
   |||||: .....: .....: .....: .....:
660 lullelulysglInlyslsasnasnlysarlysllelyrhnlslelysls 676

```

```

1242 TCCTCAATGTTATGGTTTAAGCATGTCATGATGCTGCCGTCATTAT 1291
      ::::: |||||
677 ASP.....IleIleIleSerLeuLe 683
1292 ACCATTTCGTTAAATTCGTCCAATCAGAGC..... 1335
      |||::: |||:::
683 uProtYr...ValasnThrIleValIgluArgGlnIlePheLysLeuLeuV 699
1326 ..CATAAATAATAGTGAATCATCGAAGCAATTAATAAATCCCATAT 1373
      ::::: |||||
699 alaLysAsnAsn.....AsnIleCysSerLysIleLysAsn 711
1374 ATATTCCTTCTTCTTATCAATCCTTCTTAATCAATTAATTTAGTACAAA 1423
      |||::: |||:::
712 IleGluThrCysAsnLeuAsnIleTyrAsnAsn.....ValAspPr 725
1424 TATATAGTAT.....GTAGCA.....ATAAATT 1446
      ::::: |||||
725 ovalValIleTyrLysAsnLysLeuAlaValIgluLysMetGluLysAsnAsn 742
1447 AC.....TGTAATCAATAGT..... 1463
      |||::: |||:::
742 TyrAspLysAsnThrCysSerIleLeuSerSerTyrLysAsnTyrLeuAsn 758
1464 .....GATATTTTCATCAAAACTTCTCTATCGCTTTTATAGC 1501
      ::::: |||||
759 IleCysAsnAspAsnThrTyrValAlaHisSerSerIle...TyrCysIL 774
1502 TTCGAAAAGCTTATTCATTTATTCAGTATCTTTATATGCAATCATG 1551
      ::::: |||||
774 eglLysAsnLeuSerHisLeuAsnIleTyrTyrGlnHisLysIleV 791
1552 TAAATGTTTCATCAATAGCCATGAAATAGTTCGTTGTTATTCATCA 1601
      |||::: |||:::
791 alaPheIleLysMet.....PheTyrIleLeuThrPheLeu 802
1602 TTATCAACTGCTCTATTTTATTCATCAATGTTATTCATTTGATTAATAT 1651
      |||::: |||:::
803 LeuAlaMetProLysLysLysTyrIleAspLeuIleIlePheSerAsnIL 819
1652 CACAAT.....TATACCTTGATTTGCCAA 1677
      |||::: |||:::
819 eIleAsnAlaLeuSerLysMetCysTyrThrTyrGlnMetTyrValVal 836
1678 TTTTATGCGCATCATTTCTATCTGTAACAATTCATCTATTTGCAAT 1727
      ::::: |||||
836 euphe.....TyrPheValAsnLys.....ValCysGly 845
1728 ATTCGAATTAATAAGTATTTCTATTGT..... 1754
      |||::: |||:::
846 IleArgIleSerGluTyrValLeuSerLysTyrPhePheArgAsnGlyLe 862
1755 .....GAAAAAATAA 1767
      |||::: |||:::
862 uValLeuLysThrValGluGluGluLysGluGluGluLysGluLysG 879
1768 AAAAAAAAAA 1778
      ::::: |||||
879 LuLysGluLys 882
seq_name: p1r2:g71613
seq documentaton block:
hypothetical protein PFB0470w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 07-May-1999
C:Accession: G71613
R:Carder, M.J.; Teitelin, H.; Garucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
Science 202, 1126-1132, 1998
A:title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; M0ID:99021743

```

```

A:Accession: G71613
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1346 <GAR>
A:Cross-references: GB:AE001397; GB:AE001362; NID:g3845192; PID:g3845193; TIGR:PFB047
A:Experimental source: clone 3D7
A:Genetics:
A:Gene: PFB0470w

alignment_scores:
      Quality: 105.50      Length: 719
      Ratio: 0.319      Gaps: 48
Percent Similarity: 46.036      Percent Identity: 21.280

alignment block:
US-09-323-427-1 x G71613 ..

Align seg 1/1 to: G71613 from: 1 to: 1346
70 AACACACACACACACATATATACCCCATCAAGTGAGAGAGACAGC 119
|||::: |||::: |||:::
649 AsnHisGluAsnSerArgAsnLysIleAlaAsnTrpGlnAsnLysIleGl 665
120 AAGCAATCTTATGTTTCTTAAATTCGA..... 147
|||::: |||:::
665 uHisAsnAsnLeuAspAsnLysAsnAsnAsnMetAspPheAsnAsnMetL 682
148 .....ATTACTAATCTTCTGAAATGATGAT 174
|||::: |||:::
682 eLThrSerProLeuTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyr 698
|||::: |||::: |||:::
175 TCG...TCCTATGCTTCTGTAC.....TACACTTATGCAATGT 212
|||::: |||::: |||:::
699 SerLysGlyTyrCysAsnLeuPheGluLeuLeuTyrGlyTyr..... 712
213 CTATTCGATTCGCGTTGACAAATGCTGTGCAAGTGAGCCGAAATTCGA 262
|||::: |||:::
713 .....AsnSerGlyTyrAsn.....L 718
263 TGTGACCACTTCATATACAAATCATTTTAATACAGTAATGCATTCGA 312
|||::: |||::: |||:::
718 eLThrSerPheSerSerLeuSerTyrVal...ValLysTyrAsnG 733
|||::: |||::: |||:::
313 AGGACATGTTATGCGAAGAGCTTTATGATCAAGAGTGCCGTAATG 362
|||::: |||::: |||:::
733 LuGlnMetPheLeu...LysLysPheLysAspSerLysGlnSer..... 746
363 ATGAAGGTGAGCGTCAAGTTGCCG...GAATTCACCTTCATTTGATTCA 409
|||::: |||::: |||:::
747 .....GluValProHisAsnPheGluIleHisLeuAspAs 758
410 TGCATGTTGCGCGGTACACGATCTGAAATCCACGTCGATTTTGTAC 459
|||::: |||::: |||:::
758 n.....IleSerAspLysIleLeuLysIleLe.....G 768
460 AACACCTGTGTCATTTGCTTATCAATTTATGTTATGTTATGTTATGTTATG 509
|||::: |||::: |||:::
768 LuGlnAsnLeuAsnHisGluAsnMetLysTyrIleIleHisAsnLeuMet 784
510 GTG.....CATATCGAGTCAATGCTTTTAC 535
|||::: |||::: |||:::
785 IleSerLeuSerLeuCysAspIleLysTyrLeuAsnLeuTyrAlaLeuCy 801
|||::: |||::: |||:::
536 ATGCAAGCTGATTAACAGCTTAGTCACAGATTGAG.....T 573
|||::: |||::: |||:::
801 sphePheIleLeuLysGluAsnTyrTyrTyrLeuSerIleAspAsnLeuT 818
574 ATTCGAATTCACAACTGCTTTTCAACATCAAAATGTCGATTCGACG... 620
|||::: |||::: |||:::
818 TyrLeuTyrLeuGluIleLeuHisArgMetLysIleTyrAsnHisAspLe 834
621 ..TATGCCGTTATGCAATTT.....TGATGTCGCA 649

```

```

835 PheTyrSerIleMetGluTyrIleAsnThrHisValHisAlaLeuGluSue 851
836 CCAACCGGTCACACAGTTCATTTGCTATCAT..... 682
837 rGlnTyrLysMetLysIlePheLeuSerTyrAsnIlePheGlnLys 868
838 ...GGTCAGCAGTTTATCATTAATGACATTCGATTCGAACCGTTGAT 730
839 eAspAsnProValAspMetLysGluMetCys.....Asp 879
840 ACTTCTGGCGGTTGCCATTCCTGCTTTGTGATGATGTAACGGTGA 780
841 PhePheLeuSerSerAsnAsnLysIleGluLysGluAsnGlyAsnAsp 896
842 TACGTGGAAATTTCAATGCTGATGATGCTCTGTGATTAATATTG. 829
843 P.....LeuMetLeuGlyLysCysThrHisGluLysAsnLeuT 909
844 .....CTAATTAATTGGAA 844
845 rPlyLeuProThrAspIleGluIleLysGluAsnLeuIleAsnGlu 925
846 ..TATCCAACAGATTTAATGGCTGGC.....CAAGAGC 876
847 AsnPheGlnLysGluLeuLeuSerAsnAspAsnAspLysMetGluPh 942
848 TCACGTATACAAATATCGCGATC.....GATCAGACCTTTCTATC 917
849 eHisAspAsnAsnLysAsnIleIleGlyHisAspLysPhePheSerAsn 958
850 AATGCCAGATCAGTATTACA.....TTAAGAACCA 949
851 AsnAspGluAsnLysIleLysLysGluLysTyrPheAsnLeuLysAsnI 975
852 AATA.....G 954
853 uIleMetValPheLysLysIleGluLysThrGluThrLeuProCysThrL 992
854 CGAATGCTGTCAGCAC.....AATGTCAGAACACCAAG 989
855 eAsnIleTyrAspTyrIleAsnPheLeuLeuIleLeuIlePheTyrGln 1008
856 GATTGGAGCTGTTAAACAGCTGTGCCGACGAAACCTGCTCAGCT 1039
857 CysAsnAsnLysIleLysGluLysAspGluLysIleAsnLeu..... 1022
858 GCGCAACTTCCTTACTCAGAAAGAT.....CTGCAGA 1074
859 ...AsnPheLeuPheSerLysAspGluAsnValIleIleThrIleGlnA 1038
860 ACCGAGAAATATCATTTGATGTACGAA.....CTGATATCACACCC 1115
861 snglu.....MetTyrGluLysAsnAsnLysIleLysAsnPro 1050
862 TTGAATTAAGC.....ATGATTAATCAAGCTTGGCAGTT 1150
863 CysLysTyrValLysAsnLysGlnTyrMetLeuAspLysTyrSerGluMe 1067
864 GATTTAACGTCACCGCTGCACTTCTGCAACATATGCAACCTGTAATACT 1200
865 t.....LeuLysGluAsnLeu..... 1072
866 TGCTCAGTACAAATGATCTGCATGTCACATTTGGCTTCATATGT 1250
867 .....PheAsnIleGluSerSerLeuIleGlnLeuPheSer..... 1084
868 TTATGGGTTTAAGCATTCAT.....TGATTCCTGCC 1282
869 .....IlePheValAsnLeuLeuGluLysGluLysAspAspLysGluLeuPheVal 1101
870 GTCATTTATACATTTTCGTTAAATTTGTCCTCAATTCAGAGGCAATAAA 1332
871 ..... 111
872 ..... 111
873 ..... 111
874 ..... 111
875 ..... 111
876 ..... 111
877 ..... 111
878 ..... 111
879 ..... 111
880 ..... 111
881 ..... 111
882 ..... 111
883 ..... 111
884 ..... 111
885 ..... 111
886 ..... 111
887 ..... 111
888 ..... 111
889 ..... 111
890 ..... 111
891 ..... 111
892 ..... 111
893 ..... 111
894 ..... 111
895 ..... 111
896 ..... 111
897 ..... 111
898 ..... 111
899 ..... 111
900 ..... 111
901 ..... 111
902 ..... 111
903 ..... 111
904 ..... 111
905 ..... 111
906 ..... 111
907 ..... 111
908 ..... 111
909 ..... 111
910 ..... 111
911 ..... 111
912 ..... 111
913 ..... 111
914 ..... 111
915 ..... 111
916 ..... 111
917 ..... 111
918 ..... 111
919 ..... 111
920 ..... 111
921 ..... 111
922 ..... 111
923 ..... 111
924 ..... 111
925 ..... 111
926 ..... 111
927 ..... 111
928 ..... 111
929 ..... 111
930 ..... 111
931 ..... 111
932 ..... 111
933 ..... 111
934 ..... 111
935 ..... 111
936 ..... 111
937 ..... 111
938 ..... 111
939 ..... 111
940 ..... 111
941 ..... 111
942 ..... 111
943 ..... 111
944 ..... 111
945 ..... 111
946 ..... 111
947 ..... 111
948 ..... 111
949 ..... 111
950 ..... 111
951 ..... 111
952 ..... 111
953 ..... 111
954 ..... 111
955 ..... 111
956 ..... 111
957 ..... 111
958 ..... 111
959 ..... 111
960 ..... 111
961 ..... 111
962 ..... 111
963 ..... 111
964 ..... 111
965 ..... 111
966 ..... 111
967 ..... 111
968 ..... 111
969 ..... 111
970 ..... 111
971 ..... 111
972 ..... 111
973 ..... 111
974 ..... 111
975 ..... 111
976 ..... 111
977 ..... 111
978 ..... 111
979 ..... 111
980 ..... 111
981 ..... 111
982 ..... 111
983 ..... 111
984 ..... 111
985 ..... 111
986 ..... 111
987 ..... 111
988 ..... 111
989 ..... 111
990 ..... 111
991 ..... 111
992 ..... 111
993 ..... 111
994 ..... 111
995 ..... 111
996 ..... 111
997 ..... 111
998 ..... 111
999 ..... 111

```

```

1101 LasnGlnIleMetPheIleLeuAspPheIleLysIleIleLeuAsnGluLys 1118
1102 ATAAATGATAGATCATCGAACCAATA..... 1358
1103 aTyrIleAsnIleMetLysIleValLysLysMetLysAsnTyrAspGlu 1134
1104 ...ATMAACGCCATATATATATTCCTTCTTCTATTCATCCTTCAATA 1405
1105 AsnIleLysArgLysAsnTyrPheThr.....TyrSerLysAsnLys 1149
1106 CTAAATTTAGCTAACAAATATATAGTAT.....GTAG 1437
1107 sTyrPheGlnLeuLysLysIleAspLeuGluTyrIleAsnSerAsnIleA 1166
1108 GAAATATATACGTATATCAATACGATATTTTCATCA..... 1477
1109 snAsnLysLysLysAsnThrTyrAsnAspPhePheAsnGluAsnAsn 1182
1110 AACCTTCTCTATCGCTTTATAGCTTCTGMAAAGCTTATTCATAT 1524
1111 IleAsnTyrArgTyrGlnTyrGlnSerValHisLysAlaIleGlnLeuPh 1199
1112 CAGTAT.....CTTTATATGATACATATTTGTAATGT..... 1558
1113 eSerAspAsnIleIleArgTyrSerHisAsnGluLysIleAsnThrHisT 1216
1114 ...TTCATCATTAAGCCA..... 1573
1115 yLysAsnAsnLysTyrIleIleLysAspIleLysThrPheTyrLysLeu 1232
1116 TGAATAGTTTGGTTGTTATATATCATAT..... 1603
1117 AspAsnPheLeuIleSerAspIleLeuLeuIleGluLysGluAsnLys 1249
1118 ATCAACTTGTCCATTTTATCTA.....A 1628
1119 sGluGlnIlePheTyrPheLeuLeuPheTyrProPheGluLeuLysGlnT 1266
1120 CAGTTATATCATTTGATATATATCAAAATATATACCTGTATGCCAAT 1678
1121 hValIleHisIle.....LysAsnAsnThrPheLeuPheAsn 1278
1122 TTT...TATGGCATCATTTCCCTATTCGTAAACAATTCATTTATTTCA 1725
1123 TyrLysTyrAspGluThrPheLeuPheAsnMetGluIleLeuPheLeuTy 1295
1124 T 1726
1125 r 1295
1126 r 1295

```

seq_name: plr2:S55098

seq_documentation_block:

probable membrane protein YMR216c - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein YMR261.10c

C:Species: *Saccharomyces cerevisiae*

C>Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 06-Feb-1998

C:Accession: S55098

R:Dedman, K.; Brown, D.; Bowman, S. submitted to the EMBL Data Library, June 1995

A:Reference number: S55089

A:Accession: S55098

A:Molecule type: DNA

A:Residues: 1-742 <DED>

A:Cross-references: EMBL:249809; NID:9854459; PTD:9854468; MIPS:YMR216c

A:Experimental source: strain AB972

C:Genetics:

A:Gene: SGD:SKY1

A:Cross-references: SGD:S0004829; MIPS:YMR216c

A:Map position: 13R

C:Keywords: transmembrane protein

F:241-257/Domain: transmembrane #status predicted <TM>

[illegible][illegible]

seq_name: pir2:S57242

```
seq_documentation_block :
```

Caenorhabditis elegans

N;Alternate names: myosin-regulating protein

C:Species: Caenorhabditis elegans
N;Collalins: protein kinase (EC 2.7.1.-)

C:Date: 28-Oct-1995 #sequence revision 24-C

C;Accession: S57242; S07571; S06797; S57218

R; Benian, G.M.; L'Hernault, S.W.; Morris, M

submitted to the EMBL Data Library, February 1

A;Description: Additional sequence complexity with

A;Reference number: S57242

A:Accession: S5/242
A:Molecule type: DNA

A:Residues: 1-6839 <BEN1>

A;Cross-references: EMBL:L10351

A; Experimental source: var. Bristol

R; Benian, G.

submitted to the EMBL Data

A;Reference number: S075711

A: Accession: 507511
A: Molecule type: DNA

A;Molecule type: DNA
A:Residues: 792-6839 <BEN2>

A;Cross-references: EMBL:X15423; NID:q6897; PIDN:CAA33463.1; PID:q68988

A;Experimental source: var. Bristol

R; Benian, G.M.; Kiff, J

Nature 342, 45-50, 1989

A;Title: Sequence of an unusually large p

A;Reference number: S06/9/; MUID:90044042
A;Accession: S06797

A:Status: nucleic acid sequence not shown

A: Molecule type: DNA
h/baccas: nucleic acid sequence not shown

A;Residues: 806-1175;1178-1998, 'Y', 2000-3

A;Cross-references: EMBL:X15423

A; Experimental source: var. Bristol

R;Benian, G.M.; L'Hernault, S.W.; Morris, M.E.
 1984: 124 1987 1104 1993

Genetics 134, 109/-1104, 1993

A11C1: additional sequence complexity in the muscle gene, unc-22, and its encoded p


```

3712 1y.....gluSerLysProLeuGluAla 3719
1040 GCGCAACTGCTTACTACAGAAAAGATCGAGACCGGAGAAATATCAT 1089
3720 GluGluProIleIleAlaLys.....AsnGlnPhe 3729
1090 TGATGTGCAACTGATATCAACACCCCTTGAATATGACGATGATATCAAG 1139
3729 eAspValProAspProValAspLysProGluValThrAspTrpAspLys 3746
1140 CTTGCCAGTGATTTAGTCACCGCTGCACCTTCTGCACATATATGACAA 1189
3746 sp...ArgIleAspIleLysTrpAsnProThrAlaAsnAsnGlyValAla 3761
1190 CCTGTA 1195
3762 ProVal 3763
seq_name: pir2:T03099

seq_documentation_block:
mucin, submaxillary - pig
N:Alternate names: apomucin
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 20-Sep-1999
C:Accession: T03099; A40009; A28528; B29789
R:Eckhardt, A.E.; Timpte, C.S.; Deluca, A.W.; Hill, R.L.
J. Biol. Chem. 272, 33204-33210, 1997
A:Title: The complete cDNA sequence and structural polymorphism of the polypeptide chain
A:Reference number: 214839; M0ID:98070526
A:Accession: T03099
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1328 <ECK>
A:Cross-references: EMBL:AF005273; NID:92581863; PID:92581864
R:Eckhardt, A.E.; Timpte, C.S.; Abernethy, J.L.; Zhao, Y.; Hill, R.L.
J. Biol. Chem. 266, 9678-9686, 1991
A:Title: Porcine submaxillary gland contains a cysteine-rich, carboxyl-terminal domain in
A:Reference number: A40009; M0ID:91236743
A:Accession: A40009
A:Molecule type: mRNA
A:Residues: 12139-12167, 'T', 12169-13288 <ECK>
A:Cross-references: GB:M61883; NID:9454837; PID:9164374
R:Timpte, C.S.; Eckhardt, A.E.; Abernethy, J.L.; Hill, R.L.
J. Biol. Chem. 263, 1081-1088, 1988
A:Title: Porcine submaxillary gland apomucin contains tandemly repeated, identical seque
A:Reference number: A28528; M0ID:88087170
A:Accession: A28528
A:Molecule type: mRNA
A:Residues: 12139-12167, 'T', 12169-12641 <TIM>
A:Cross-references: GB:J03512
A:Experimental source: submaxillary gland
R:Eckhardt, A.E.; Timpte, C.S.; Abernethy, J.L.; Toumadje, A.; Johnson Jr., W.C.; Hill,
J. Biol. Chem. 262, 11339-11344, 1987
A:Title: Structural properties of porcine submaxillary gland apomucin.
A:Reference number: A92606; M0ID:87280230
A:Accession: B29789
A:Molecule type: protein
A:Residues: 1572-1607 <ECK>
C:Superfamily: pig submaxillary mucin
C:Keywords: tandem repeat

alignment_scores:
Quality: 104.00 Length: 405
Ratio: 0.654 Gaps: 20
Percent Similarity: 39.259 Percent Identity: 21.975

alignment_block:
us-09-323-427-1 x T03099 ..
Align seg 1/1 to: T03099 from: 1 to: 13288

```

```

23 GTGTCTATTAACACCGACTGCACCAACAACAACAACAACAACAAC 72
12907 ValThrValGlyThrThrValAlaProGlySerPheAsnThrLysAlaThr 12923
73 AACAAACAACGCAATTAATTAACCCCATCAAGTGGAGGAAGAGAGAG 122
12923 ThrProThrGluValArgAlaAlaThrGlyLysIle..... 12935
123 CAATCTTGATTCTTCAAAAATCGAATTACTAATCTTCTGAATGATG 172
12935 ..... 12935
173 ATTCGCTTATTGCTTCTCTGACTACACTTAATGCAATGCTTATTCGAT 222
12936 .....ThrAlaValGlyAlaThrIleSerArgSerThr 12945
223 TCCGGTTGCAATGGTGTGCA.....GCTGAGCCAGAAATG 260
12945 rGlyIleSerThrGlyProGluAsnSerThrProGlyThrThrGluThrG 12962
261 AATGTGACCAACTTCATTAACATCAATTTTAATACAGT...AATGCA 307
12962 LysSerGlyThrThrSerSerProGlyGlyValLysThrGluAlaThrThr 12978
308 TTGCAAGACATGTTTATGTGAAGGCTTTATGATCAAGAAGTTGCCG 357
12979 PheLysGly..... 12981
358 TAATATGATAGGTGACGTCAGTCCGGAATTCA.....CTTC 398
12982 .....ValGlyThrThrGluAlaGlyIleSerSerGlyAsnSerP 12995
399 CATTGATTCATGATGCAATGTCGCGGTACACAGTCTGTGAATCCAGTGGT 448
12995 ro...GlySerGlyValThrSerSerGlnGlyIleThrSerArgGlu 13010
449 ATTTTGTGAACAACAAGCTGTGTCATTTGTTTCATTCATTAATTTGTAAC 498
13011 AlaSerGluThrThrThr..... 13016
499 CAAAGTTGATGTCGATATCGAGTACATGCTTT...TACATGGAAGCTG 545
13017 .....AlaProArgIleSerAlaThrGlySerThrSerValS 13029
546 ATAAACAGTTAGTGCACAGATGAGTATGAA...ATCACACTGCT 592
13029 erLysGluIleThrAlaSerProLysValSerSerProGluThrThrAla 13045
593 TTCAACTCAA..... 604
13046 GlyAlaThrGluAspGlnGluAsnGluAsnLysThrGlyCysProAlaPr 13062
605 .ATGTCCCGATGCCAGATATGCCGTTATGAATTTTGGATGGTGGACCAA 653
13062 OeuProProProProValCysHis.....GlyProL 13073
654 CCGGTCAACCAAGTCAATTTCTATCATGTGCAGCAGTATATCAATAA 703
13073 euGlyGluGluLysSerProGlyAspValThrThrAlaAsnLysHisLys 13089
704 TGGACATGC...GATTCGAAACCGTTGAT.....ACTTCTGCCG 741
13090 CysThrCysThrGluAlaLysThrValAspCysLysProLysGluCysPr 13106
742 GGTGTCCATTCCTGCTTGTGATGAT.....GGTA 773
13106 oSerProProThrCysLysThrGlyGluArgLeuIleLysPheLysAlaA 13123
774 ACGGTGATACGTGTGAATTTGAATGCTGATGATGCTGTGATATAA 823
13123 snAspThrCysCysGluIle.....GlyHisCysGluLysArgThr 13136
824 TATTTGCTAATAATTTGGAATATCCAAACAGATTAATGCGTGGCCAGA 873

```

```
13137 cysleupheasnasnThrAspTyr..... 13144
      ||:::||||| ::|||
874 AGCTCAGTATACAAATATGCCGATGCATCACAGCTTTCTATCAATGCC 923
13145 .....G 13145
924 AGATCAGTATTACCATTAAGAACAATAGCGAATGTGTCAGCCACAA 973
      ::::: ::|||::: |||||
13145 luValGlySerSerPheAspProAsnAsnProCysValThrTyrSer 13161
974 TGTTCAGAACCAACAGATTGGAGCTTTAA.....AA 1007
      || ::: ||||| |||||
13162 Cys...GlnAsnThrGlyPheThrAlaValAlGlnAsnCysProLysGI 13177
1008 CAGGTGTGCCGC.....AGCAAAACCTGCTGC. 1035
      : |||||
13177 nThrTyrCysAlaGluGluAspArgValTyrAspSerLysGlnCysCysT 13194
1036 ..AGCTGCCCAACTTCGTTACTCAAGAAAGATCTGACAGAACCGGAGAA 1083
      ::||| ::||| ::: |||
13194 yrThrCysLysSerSerCysLysProSerProValAsnValThrValArg 13210
1084 TATCATGTGATGACG 1098
      ||::: |||||
13211 TyrAsnGlyCysThr 13215
```


alignment_scores: Quality: 1420.50 Length: 416
 Ratio: 4.228 Gaps: 7
 Percent Similarity: 80.769 Percent Identity: 67.067

alignment_block:
 US-09-323-427-1 x CUT1_CAEEL ..

Align seg 1/1 to: CUT1_CAEEL from: 1 to: 424

```

191 TGTACACACTATATTCGATTCCTTATTCG...ATTCGGTTGACAATGG 237
   ||| : : : : : ||| ||| ||| ||| ||| : : : : :
8 CysLeuAlaLeuAlaLeuSerAlaSerAlaLeuProValAspAsnAs 24

238 TGTGAGGTGAGCCGCAAAATGATGTGACCACTTCATTAACATCA 287
   : : : : : ||| ||| ||| ||| ||| : : : : :
24 nValIgluIgluProIgluValIgluIgluIgluIgluIgluValA 41

288 ATTTTATATACAGTAATGCATTGGAAGACATGTATGTGAAGCTTT 337
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
41 snPheAsnThrArgAsnProPheIgluIgluIgluIgluIgluVal 57

338 TATGATCAAGAAGGTGGCCGTAATGATGAGGAGGAGTCAAGTTCGG 387
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
58 TyrAspGlnAlaIgluIgluIgluIgluIgluIgluIgluIgluIglu 74

388 AATTCACATTCATTTGATTCATGCAATGTGGCGGTACACAGATCTGTA 437
   ||| : : : : : ||| ||| ||| ||| ||| ||| ||| ||| |||
74 yIleGluLeuProPheAspSerCysAsnThrAlaArgThrArgSerLeu 91

438 ATCCAGCTGATATTTTGTACACAACTGTCTCATTTCTGTTCTATCCA 487
   ||| : : : : : ||| ||| ||| ||| ||| ||| ||| ||| |||
91 snProIgluIgluValPheValSerThrThrValValIleSerPheHisPro 107

488 TTTATTTACCAAAAGTTGATGTCATATCGATATCGAATGCTTTACAT 537
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
108 GlnPheValThrIgluValAspArgAlaIgluIgluIgluIgluIgluIglu 124

538 GGAAGCTGATAAACAGTAGTGCACAGATGAGATGTGAAATGACAA 587
   ||| : : : : : ||| ||| ||| ||| ||| ||| ||| ||| |||
124 tGluSerAspIgluIgluIgluIgluIgluIgluIgluIgluIgluIglu 141

588 CTGCTTTTAAACTGAAATGTCGCCGATGCGCATATCCGCTTATGAAAT 637
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
141 hValAlaPheIgluIgluIgluIgluIgluIgluIgluIgluIgluIglu 157

638 TTGATGATGAGCAACCGGTACACAGTTCATTTGATTCATTTGATCA 687
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
158 LeuAspGluIgluIgluIgluIgluIgluIgluIgluIgluIgluIglu 174

688 GCCAGTTTATCATAAATGACATGCGATTCTGAAACCGTTGATTTCTCT 737
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
174 nGlnIgluIgluIgluIgluIgluIgluIgluIgluIgluIgluIgluIglu 191

738 GGGCGGTTTCCATTCCTGCTTTGATGATGATGATGATGATGATGATG 787
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
191 yAlaIgluIgluIgluIgluIgluIgluIgluIgluIgluIgluIgluIglu 207

788 GAAATTCCTAAATGCTGATGATGCTGCTGATTAATATTTGCTAAATA 837
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
208 GlnIleLeuAsnGluIgluIgluIgluIgluIgluIgluIgluIgluIglu 224

838 TTTGGAATATCCACAGATTTAATGGCTGGCCAGAGAGCTCACGATATCA 887
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
224 nLeuGluIgluIgluIgluIgluIgluIgluIgluIgluIgluIgluIglu 241

888 AATATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 937
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
241 ySTyTAlaAspArgSerGlnLeuPheTyrGlnCysGlnIleSerIleThr 257

938 ATTTAAGAACCAATGAGATGTGTTGACCAACGATGTTGAGAACCAACA 987
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
258 IleIysAspProIgluIgluIgluIgluIgluIgluIgluIgluIgluIglu 274

```

```

988 AGGATTCGAGCTGTAAACAGCTGCT..... 1015
274 nGluPheGluIgluIgluIgluIgluIgluIgluIgluIgluIgluIgluIglu 291
1016 .....GCCGCA 1021
291 LaAlaProGlnAlaIgluIgluIgluIgluIgluIgluIgluIgluIgluIglu 307
1022 GCAAACCTGCTGCACCT..... 1039
308 AlaAlaProValAlaAlaProValAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 324
1040 .....GCCCACTTCGTTTACTC...AAGAAAGATCT. 1069
324 lProAlgalAlaThrLeuAlaIgluIgluIgluIgluIgluIgluIgluIgluIglu 341
1070 ..GCAGAACCGGAGATATCATTTGATGTACGAACTGATTCACAAACCTTT 1117
341 hGluIgluIgluIgluIgluIgluIgluIgluIgluIgluIgluIgluIgluIglu 357
1118 GAATATGCGATGATGATATCAAGCTTGGCCAGTTGATTTACGTACACCGTGC 1167
358 AspIleMetGluGlu.....AlaSerProSerAlaProGluAlaAlaAla 372
1168 ACTTCTGCACATATGATGACAACTGTAATTAATCTGCTGCTGACGATCAAAATG 1217
372 aLeuValSerGlu.....GluSerValArgArgArgAlaThrSerThrG 387
1218 GAATCTGCATGACCATTTGGCTCTCAATGTTATGATGTTAAGCAT 1267
387 lYleCysLeuThrProIleGlyPheAlaSerPheLeuIgluIgluIgluIgluIglu 403
1268 GCATTGATTCGCTGCCGCTCATTTATACCATTTGCTTAAATTCGTCCA 1315
404 lIleValAlaThrAlaLeuSerAlaThrIlePheTyrValAlaArgPro 419

seq_name: SwissProt_38:RPOB_PLAFA

seq_documentation_block:
ID: RPOB_PLAFA       STANDARD;       PRT; 1024 AA.
AC P21421;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE BETA CHAIN (EC 2.7.7.6).
GN RPOB.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BW(C10);
RC MEDLINE: 95107345.
RA GARDNER M.J., GOLDMAN N., BARNETT P., MOORE P.W., RANGACHARI K.,
RA STRATH M., WHITE A., WILLIAMSON D.H., WILSON R.J.M.:
RT "Phylogenetic analysis of the rpoB gene from the plasid-like DNA of
RT Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 66:221-231(1994).
RN [2]
RP SEQUENCE OF 328-1024 FROM N.A.
RX MEDLINE: 91187055.
RA GARDNER M.J., WILLIAMSON D.H., WILSON R.J.M.:
RT "A circular DNA in malaria parasites encodes an RNA polymerase like
RT that of prokaryotes and chloroplasts.";
RL Mol. Biochem. Parasitol. 44:115-124(1991).
CC -I- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -I- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC RNA(N).
CC -I- MISCELLANEOUS: THIS RNA POLYMERASE IS ENCODED ON A CIRCULAR DNA.
CC -I- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC

```



```

489 snleuwtserileysmethisthrglnllevalproilleletyrpro 505
1377 TTCGTTTCCTTATCATCCTCTTAATACTAATTATGCTACAATAT 1426
      :::::
506 AsnleuSerAsnle.....IleThrAsnTy 514
1427 ATAGATATGAGAAATATTAATCTGTATATACATA.....A 1461
      ::::: ||::: ||:::
514 rAsnPhelIleLeuAsnLysTyrLeuAsnHisLeuIleIleSerTyrGlnG 531
1462 GTGATATTTTCATCAAAACTTCTTATGCGCTTTTATAGCTTGAAG 1511
      :: ||| ||| :::: |||
531 IuGlyIleValIleTyrValSerCysIleIleIleIleArgasp... 546
1512 CTATATCATATATTCAGTAATCTTTATATGACATATTT..... 1550
      ||::: ||::: ||::: ||::: ||::: ||:::
547 LeuPheAsnArgIleIleIleTyrTyrLeuAsnAsnTyrLysIleAs 563
1551 .....G 1551
563 ngIAsnIleLeuIleTyrLysProIleValTyrValGlyLulysV 580
1552 TAAATGTTTCATCATAGGCCATGATAGTTGCTTTGTTATATCATCA 1601
      ||::: ||::: ||::: ||::: ||::: ||::: ||:::
580 alAsnIleGlyIleIleLeuAlaIleAsnSerAsnleuAsnSerGlu 596
1602 TTATCAACTTTCCTATTTTATCTAACAGTTT..... 1634
      ||| ::::
597 TyrSerLeuGlyAsnAsnleuValGlyTyrGlySerTyrLeuGlyTy 613
1635 .....ATCATTTGTGATTAATATACACAATTAATACCT 1665
      ||::: ||::: ||::: ||::: ||::: ||::: ||:::
613 rgluTyrGluAspAlaIleIleIleSerArgLysIleLeuTyrAsnAsn 630
1666 TGTATGCCCCAATTTTATGGGCATCATTTCTATCTGTAAAC..... 1709
      ||::: ||::: ||::: ||::: ||::: ||::: ||:::
630 euTyrThrSerLeuHisLeuAsnIleTyrGluIleSerLeuAsnIleIle 646
1710 AATTCACCTT.....ATTGCATATATGCAATTAAGATATTTTC 1748
      ||::: ||::: ||::: ||::: ||::: ||::: ||:::
647 AsnAsnIleProGluIleCysSerIleAsnleuSerLysMetTyr 661
seq_name: SwissProt_38:YOH3_CAEEL
seq_documentation_block:
ID YOH3_CAEEL STANDARD: PRT: 495 AA.
AC Q09276;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEtical 55.1 KD PROTEIN C43C3.3 IN CHROMOSOME X.
GN C43C3.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA SU1STON J.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z47067; CAAB7330.1; -
CC WORMPEP: C43C3.3; CE01525.
CC Hypothetical protein; Transmembrane; ATP-binding.
KW TRANSMEM 11 31 POTENTIAL.

```

```

FT TRANSMEM 426 446 POTENTIAL.
FT NP_BIND 171 178 ATP (POTENTIAL).
SQ SEQUENCE 495 AA: 55068 MM: 09C77468 CRC32;

alignment_scores:
      Quality: 109.50      Length: 415
      Ratio: 0.559      Gaps: 24
Percent Similarity: 47.229      Percent Identity: 21.446

alignment_block:
US-09-323-427-1 x YOH3_CAEEL ..
Align seg 1/1 to: YOH3_CAEEL from: 1 to: 495

35 CGAGCTGGCAGCAACAACACAAACAAACAAACAAACAAACAAACAGC 84
   |||
   ::::: |||
26 ProIleuValMetCpHePheAsnAsnGlyAlaAsnThrSerLeuIleH 42
   ::::: |||
85 AATATATACCCCATCATAGTGA.....GGAGAGACAGAG 119
   ::::: |||
42 spHeLeuGlyProHisSerProSerSerLeuPheLeuAlaHisThrA 59
120 AAGCAATCTTAACTTTTCTAAAAATCGAA.....TTTACTAAATCT 160
   :: ::::: |||
59 spHisLeuLeuAlaTyrAlaThrSerGluAlaGlyAlaPheValArg 75
161 TCGAATATGATGATTCGTCTTATGCTTCTGTACTACACTAATGCAAT 210
   ::::: |||
76 MetArgAlaLysLeuGluIleuVal.....AspCysIleAlaAs 88
211 GCTTATATGATTCGCGTTGACAAATGCTGTCGAGGTGACGAGAAATG 260
   ||::: ||::: ||::: ||::: ||::: ||::: ||:::
88 pserPheThrValIleuAsn.....LysSerAspProGluValM 102
261 AA.....TGTGACCAACTTCATATACAAATTAATATACAGCT 301
   ::::: |||
102 eArgMetIleSerAsnProLysSerGlnProVal..... 113
302 AATGCAATTCGAGACAGCATGTTATGTAAGAGCTTTATGATCAGAAG 351
   ||::: ||::: ||::: ||::: ||::: ||::: ||:::
114 .....ValTyrValTyrGlyHisLysThrArgHisP 124
352 TTGC.....CGTATGATGAAGGTGACGTCAAGTTGCCGGAATTT 392
   |||
   ::::: |||
124 ocYsgIlyThrSerMetLysAspGluLysGlyLeuThrAsnPheAsnLeu 141
393 CACTGCCATTTGATTC...TGCATGTTGCCGTCACAGATCTGTAAT 439
   ::::: ||::: ||::: ||::: ||::: ||::: ||:::
141 hTrIleProTyrGlySerGluCysAspValThrLeuThrAspLeu..... 155
440 CCAGCTGATTTTGTAAACACACAGCTGTCATTTGCTTGCAT...CC 486
   ||::: ||::: ||::: ||::: ||::: ||::: ||:::
156 ProLysHisArgTyrAlaGluThrThrValValLeuGluAspAsnAlaAs 172
487 ATTATTTGTTACCAAGTTGATCGCATATGCAATGCAATGCTTTTACA 536
   |||
   ||::: |||
172 pleuSerPheGlyLysThrThrArgLeuAsnHisValPheCysLeuTyr 189
537 TGGAGCGTGATTAACACAGTACTGCACAGATTGAGGTATTCGAATTCAC 586
   ::|
   ||::: |||
189 hArgAsnValLysThr.....IleArgPheSerAspValSer 201
587 ACTGGTTTTCAA.....ACTCAATTCGCCGATGCCGATG 624
202 AsnGlyHisGluValIleAlaSerThrGlyLysProLysPro..... 216
625 CCGTATGAATATTTG.....GATGCTGACCAACCGGTCAAC 662
   ::|
   ||::: |||
217 LysValGluMetLeuPheArgSerThrAspSerCylLysThrLeuGlnA 233
663 CA.....GTTCAATTTGCTATCATCATGTCAGGCA... 691
   ||::: |||

```

```

233 laalaargluasnlguphevalguphepellealeuserProasp 249
692 ...GTTTACATTAATGACA.....TGGATTGCA 720
250 SeratiyrhisglyllesePrrolyseclucystrpheseraspargl 266
721 AACCGTGTAT.....ACTTTCGCGGGTTGTC 749
266 uaspleseralaproaspalalylysilethrphe.....ValgIng 281
750 ATTCCTGCTTTGTGATGATGTAACGCTGATCTGTGAATTTAAAT 799
281 lylgicyspProvalasnlgymetasnasp..... 290
800 GCTGATGATGTCCTGTGATAAATATTGCAATAATTGGATATCC 849
291 .....llelleaspProleualasnva 298
850 AACAGATTAATGCTGGCCAGACGCTACAGTATACAAATGCGGATC 899
298 lasnaspInlelyrpheserlysphearGthrphearpheeGlyasng 315
900 GATCAGACCTTTTATATCAATGCCAGATCAGTATTACATTAAAGACCA 949
315 InsetThvalpheyahiscysgInvalgInvalCysleuylslys... 330
950 AATACGGAATGCTGACACCAATGTTGACAGACCAACCAAGATTGGAGC 999
331 ...AspGlucysSerlys...Thrcystyrlylslys..... 340
1000 TGTTAAMAACAGGTGGTGGCCAGACAAACCTGCTGACAGCTGCCAATTC 1049
341 .....ValserAspserAsnleuthralagInvalGlyleua 352
1050 GTTACTACAGAAAGATCTGCAGACCGGAGAAATATCATGATGTACGA 1099
352 ryphearghislysarGser.....llethraspleuInu 363
1100 ACTGATATCAACACCCCTGAATATAGCATGATATCAAGCTTG 1144
364 ArgatgthrThrargSerAlaProthrAspaspasnglySerleu 378

seq_name: SwissProt_38:PGSL_BOVIN
seq_documentation_block:
ID PGSL_BOVIN STANDARD: PRT: 369 AA.
AC P21809; P79259;
DT 01-MAY-1991 (Rel. 18, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE BONE/CARTILAGE PROTEOGLYCAN I PRECURSOR (BIGLYCAN) (LEUCINE-RICH PG I)
DE (PG-51).
DE BGN.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-AORTA;
RX MEDLINE: 96113563.
RA XU J.H., RADHAKRISHNAMURTHY B., SRINIVASAN S.R., BERENSON G.S.;
RT "Primary structure of bovine aorta biglycan core protein deduced from
RT cloned cDNA.";
RL Biochem. Mol. Biol. Int. 37:263-272(1995).

[2]
RP SEQUENCE OF 38-369.
RC TISSUE-CARTILAGE;
RX MEDLINE: 89255324.
RA NEAME P.J., CHOI H.U., ROSENBERG L.C.;
RT "The primary structure of the core protein of the small, leucine-rich
RT proteoglycan (pe I) from bovine articular cartilage.";
RL J. Biol. Chem. 264:8653-8661(1989).

```

```

RN [3]
RP SEQUENCE OF 38-63.
RC TISSUE-CARTILAGE;
RX MEDLINE: 89123388.
RA CHOI H.U., JOHNSON T.L., PAL S., TANG L.H., ROSENBERG L.C.,
RA NEAME P.J.;
RT "Characterization of the dermatan sulfate proteoglycans, DS-PGI and
RT DS-PGII, from bovine articular cartilage and skin isolated by occyl-
RT sepharose chromatography.";
RL J. Biol. Chem. 264:2876-2884(1989).

CC -1- TISSUE SPECIFICITY: FOUND IN THE EXTRACELLULAR MATRICES OF SEVERAL
CC CONNECTIVE TISSUES, SPECIALLY IN ARTICULAR CARTILAGES.
CC -1- PTM: THE TWO GLYCOSAMINOGLYCAN CHAINS ATTACHED TO BIGLYCAN CAN BE
CC EITHER CHONDROITIN SULFATE OR DERMATAN SULFATE.
CC -1- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC FAMILY.
CC -1- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC MANY PROTEINS. NUMBER IN THIS PROTEIN: 10.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S82652; ABA46746.1; -.
DR PIR: A33701; A33701.
DR PFM: PF00560; LRR; 8.
DR PFM: PF01462; LRRNT; 1.
KW Glycoprotein; Connective tissue; Extracellular matrix; Proteoglycan;
KW Signal; Repeat; Leucine-repeat.
FT SIGNAL 1 19
FT PROPEP 20 37
FT CHAIN 38 369
FT DOMAIN 93 316
FT REPEAT 93 106
FT REPEAT 117 130
FT REPEAT 141 154
FT REPEAT 162 175
FT REPEAT 186 199
FT REPEAT 211 224
FT REPEAT 232 245
FT REPEAT 256 269
FT REPEAT 280 288
FT REPEAT 303 316
FT CARBOHYD 42 42
FT CARBOHYD 48 48
FT CARBOHYD 271 271
FT CARBOHYD 312 312
FT DISULFID 64 77
FT DISULFID 322 355
FT CONFLICT 152 152
FT CONFLICT 188 188
FT CONFLICT 354 354
FT CONFLICT 368 369
SQ SEQUENCE 369 AA; 41509 MM; FIC673B CRC32;

alignment_scores:
Quality: 106.50 Length: 299
Ratio: 0.873 Gaps: 13
Percent Similarity: 40.803 Percent Identity: 20.736

alignment_block:
US-09-323-427-1 x PGSL_BOVIN ..
Align seg 1/1 to: PGSL_BOVIN from: 1 to: 369

1005 AACAGGTGGTGGCCAGACAAACCTGCTGACAGCTGCGCACTTGCTTTA 1054
|||||: |||: :|||:|||||:|||||
87 lylgIuileSerProaspPrthrThrleuasnaspInuAsnAsnAspDI 103

```

```

1055 CTGA.....AGAAAAGAT..... 1067
103 eSeGluLeuArgLysAspRheLysGluGlnHisLeuTyrAlaL 120
1068 ....CTGCAGAACCGAGAAATATCATTTAGTACGACAGATATCAACA 1112
120 euValLeuValAsnLysIleSerLysIleGlnGlySAlaPheSer 136
1113 CCCTGAAATTAAGGATGATATCAACCTTTGCCAGTTGATTTACGTCAC 1162
137 ProLeuTyr.....LysLeuGlnLysLeuTyrIleSerLys 148
1163 CGTGCACCTTCTGCACATATATGACACACCTGATACCTCTCGACATACA 1212
148 sAsnHisLeuCysGluIleProProAsn..... 157
1213 AATGGAATCTGCATCTACCATTTGGCTTCTCANATTTATGGGTTAA 1262
157 ..... 157
1263 GCATTCGATGATTGCTGCCGTCATTTATACATTTGCTTAATTTCGT 1312
158 .....LeuProSerSerLeu.....Val 163
1313 CCAATATCAGAGGATATAAATATATGTTAGA.....ATCA 1347
163 IGluleuArgIleHisAspAsnArgIleArgLysValProLysGlyValP 180
1348 TCGACAGCAAA..... 1358
180 heSerGlyLeuArgAsnMetAsnCysIleGluMetGlyLysProLeu 196
1359 .....ATAAAGTCGCATA 1372
197 GluAsnSerLysPheGluProGlyAlaPheAspGlyLeuLysLeuAsnTy 213
1373 TATATTCGTTCTTCTTATCATCTCTTAATATTAATTTAGTACAA 1422
213 rLeuArgIleSerGluAlaLys.....LeuThrG 223
1423 ATATATAGTATAGTACGAAATATATCTGATATACATATGATATTTTC 1472
223 LysLe.....ProLysAspLeuProGluThrLeuAsnGluLeuHis 236
1473 ATCAAAATCTCTTATCATCGCTTTATAGCTTCTGAAAGCTTATTCATA 1522
237 LeuAspHisAsnLysIleGlnAlaIleGluLeuGlnAspLeuAlaGly 253
1523 TTCAGTAATCTTTATATATGCATATGTAATGTTTCATCATTAAGGCC 1572
253 rSerLysLeuTyrArgLeu.....GlyLeuGln 263
1573 ATGAATATAGTTCTGTTGTTATATCA...TCATATATCACTTGCTCT... 1616
263 IAsnGlnIleArgMetIleGluAsnGlySerLeuSerPheLeuProThr 279
1616 ..... 1616
280 LeuArgGluLeuHisLeuAspAsnLysLeuSerArgValProAlaG 286
1617 .....ATTTTATCTAACAAGTTTATCATTTGATATAATA 1650
296 yLeuProAspLeuLysLeuLeuGlnValValTyrLeuHisThrAsnAsn 313
1651 TCACAATTTATACCTTG.....TATTTGCCAATTTTATAGGCATC 1691
313 leuTyrLysValGlyValAsnAspPheCysProValGlyPheGlyVal 328
seq_name: SwissProt_38:KM65_YEAST
seq_documentation_block:
ID KM65_YEAST STANDARD: PRT: 742 AA.

```

```

AC Q03656;
DR 01-NOV-1997 (Rel. 35, Created)
DR 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PROBABLE SERINE/THREONINE-PROTEIN KINASE YMR216C (EC 2.7.1.-).
GN YMR216C OR YMR261.10C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972.
RA DEDMAN K., BROWN D., BOWMAN S., BARRELL B.G., RAJANDREAM M.A.,
RA WALSH S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -! SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
CC PROTEIN KINASES. SIMILAR TO S.POMBE DSK1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
CC EMBL: 249809; CAA89931.1; -.
CC DR HSSP: P24941; IAO1.
CC DR PROSITE: PS00107; PROTEIN_KINASE_ATP_1.
CC DR PROSITE: PS00108; PROTEIN_KINASE_ST_1.
CC DR PROSITE: PS50011; PROTEIN_KINASE_DOM_1.
CC DR PFM: PF00069; pkinase; 3.
CC KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;
CC ATP-binding.
CC FT DOMAIN 158 706 PROTEIN KINASE.
CC FT NP_BIND 164 172 ATP (BY SIMILARITY).
CC FT BINDING 187 187 ATP (BY SIMILARITY).
CC FT ACT_SITE 294 294 BY SIMILARITY.
CC FT SEQUENCE 742 AA; 83237 MW; DCF32417 CRC32;

```

```

alignment_scores:
  Quality: 104.50      Length: 303
  Ratio: 0.816        Gaps: 13
  Percent Similarity: 42.244  Percent Identity: 20.462

```

alignment_block:

US-09-323-427-1 x KM65_YEAST ..

Align seg 1/1 to: KM65_YEAST from: 1 to: 742

```

43 AGCAACAACAACAACAACAACAACAACAACAACAACAATTAATA 92
   |||:|||||:|||||:|||||:|||||:|||||:|||||
419 SerGlyAsnArgAsnIleProSerSerIleAsnAsnSerIleAsn 435
   | |||:|||||:|||||:|||||:|||||:|||||:
93 CCCCATCAAGTGGAGAGACAGACAGACAGACATCTTTTCTTAATA 142
   | |||:|||||:|||||:|||||:|||||:|||||:
435 nGlyIleGlyIleLysAsnSerAsnAsnSerPheLeuAsnSerValProH 452
   | |||:|||||:|||||:|||||:|||||:|||||:
143 ATGCAATTTACTAATATCTTCTGAATGATGATTCGTTTATGGCTTCTG 192
   : : |||:|||||:|||||:|||||:|||||:
452 IserValThrArgMetPheIleAsnGluAspSerAsn..... 464
   |||:|||||:|||||:|||||:|||||:
193 TACTACACTTATTCGTTCTTATTCGATTCGCCGTTGCAATGGTGTCG 242
   |||:|||||:|||||:|||||:|||||:
465 .....AspAsnAsn..... 467
243 AAGGTGACGACGAATTTGAATGTGACCAACTTCAATATCAATTAATTTT 292
   : |||:|||||:|||||:|||||:|||||:
468 .....AsnAsnAspAsnSerLysAsnLysAsnAsnAsn..... 480
293 AATACACGTAATGCATTTGCAAGACATGTTTATGTGAAGAGCTTTATGCA 342
   |||:|||||:|||||:|||||:|||||:

```



```

481 AaSerAsnAsnAsnAsnGluSpliMeLanThProLeuHisG1 497
343 TCAAGAGAGTCCGATGATGAAGTGGAGTCCAGTTGCCGCAATT 392
497 uGU.....GlnLeuAlaAspSerL 504
393 CACTTCATTTGATTCATCAATGTTGCCGTAACAGATCTCGAATCA 442
504 euSerThrPheAspIleSerAsnIleSerGlnSerAspThrAsn... 519
443 CGTGGATTTTGTACACACACTGTTCATTTCTTCATTCATTAATT 492
520 ..GlyProTyrIleSerAsnTh..... 526
493 TGTTCACCAAGTTGATCGCATATGAGTACATGCTTTACATGAG 542
527 .....MetAspS 529
543 CTGATTAACAGTTAGTCA.....CAG 565
529 eAsnSerAsnValSerThrAspIleAsnSerProGluAsnLeuIleGln 545
566 ATTGAGTATCTGAATC.....AC 585
546 IleLysIleAlaAspLeuGlyAsnAlaCysTyrPyrAspGluHisTyrTh 562
586 AACTGCTTTTCAACACTCAATGTCGCCGATGCCAGATGCCGTTATGAA 635
562 rAsnSerIleGlnThrArgGluTyrArgSerPro.....GluV 575
636 TTTTGGATGGTGACCAACCGGTCAACAGTTCAATTGCTATCATTTGGT 685
575 alLeuLeuGlyAlaPro..... 580
686 CAGCCAGTTTATCATTAATGACATGCGATTCGAAACCGTTGATACTTT 735
581 .....TrpGlyCysGlyAlaAspIleTrpSerThrAl 591
736 CTGCGCGGTGTC..... 748
591 acySLeuIlePheGlnLeuIleThrGlyAspPheLeuPheGlnProAsp 608
749 .....CATTCCTGCTTTGTGCATGATGTTAAC...GGTGAATCTGTGAA 790
608 lUGlyHisSerTyrThrLysAspAspHisIleAlaGlnIleIleGlu 624
791 ATTCTAAAGCTGATGATGCTGCTTGAATAATATTTGCTAAATAATT 840
625 LeuLeuGly.....GluLeuProSerTyrLeuLeuArgAsnG1 637
841 GGAATAT 847
637 yLysTyr 639

seq_name: swissprot_38:APMU_PIG
seq_documentation_block:
ID_APMU_PIG STANDARD: PRT: 1150 AA.
AC P12021:
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE APOMUCIN (MUCIN CORE PROTEIN) (FRAGMENT).
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheraia; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SUBMAXILLARY GLAND;
RX MEDLINE: 91236743.
RA ECKHARDT A.E., TIMPTE C.S., ABERNETHY J.L., ZHANG Y., HILL R.L.;
RT "Porcine submaxillary mucin contains a cysteine-rich,
carboxyl-terminal domain in addition to a highly repetitive,
```

```

RT glycosylated domain.";
RL J. Biol. Chem. 266:9678-9686(1991).
RN [2]
RN SEQUENCE OF 1-503 FROM N.A.
RP TISSUE-SUBMAXILLARY GLAND;
RC MEDLINE: 88087170.
RA TIMPTE C.S., ECKHARDT A.E., ABERNETHY J.L., HILL R.L.;
RT "Porcine submaxillary gland apomucin contains tandemly repeated,
identical sequences of 81 residues.";
RL J. Biol. Chem. 263:1081-1088(1988).
RN [3]
RP SEQUENCE OF 45-80.
RC TISSUE-SUBMAXILLARY GLAND;
RX MEDLINE: 87280230.
RA ECKHARDT A.E., TIMPTE C.S., ABERNETHY J.L., TOUMADJE A.,
JOHNSON W.C. JR., HILL R.L.;
RT "Structural properties of porcine submaxillary gland apomucin.";
RL J. Biol. Chem. 262:11339-11344(1987).
RN [4]
RP CARBOHYDRATE-BINDING SITES, AND SEQUENCE OF 45-125.
RC TISSUE-SUBMAXILLARY GLAND;
RX MEDLINE: 97248516.
RA GERKEN T.A., OWENS C.L., PASUMARTHY M.;
RT "Determination of the site-specific O-glycosylation pattern of the
porcine submaxillary mucin tandem repeat glycopeptide. Model proposed
for the polypeptide:galnac transferase peptide binding site.";
RL J. Biol. Chem. 272:9709-9719(1997).
CC -1- FUNCTION: APOMUCIN IS PART OF MUCIN, THE MAJOR GLYCOPROTEIN
GLAND. ITS HIGHLY VISCOUS AQUEOUS SOLUTIONS SERVE TO LUBRICATE
THE ORAL CAVITY AND TO PROTECT IT FROM THE EXTERNAL
ENVIRONMENT.
CC -1- SUBUNIT: INTERMOLECULAR DISULFIDE BONDS COULD HELP MAINTAIN A
MULTIMERIC MUCIN STRUCTURE.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: SUBMAXILLARY MUCOSAE.
CC -1- DOMAIN: CONTAINS TANDEMLY REPEATED, IDENTICAL SEQUENCES OF 81
RESIDUES.
CC -1- PTM: EXTENSIVELY O-LINKED GLYCOSYLATED ON SER AND THR RESIDUES OF
THE REPEAT UNITS. HIGHEST GLYCOSYLATION APPEARS TO OCCUR ON SER
RESIDUES WHICH HAVE GLY AT POSITIONS AT +2 OR -2 FROM THE
GLYCOSYLATION SITE OR, WHERE GLY IS THE PENULTIMATE RESIDUE, THE
PRESENCE OF PROLINE (USUALLY AT POSITION +3 OR -3) APPEARS TO ALSO
ENHANCE GLYCOSYLATION.
CC -1- SIMILARITY: CONTAINS 1 WMFC DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M61883; AAA30998.1; -.
DR EMBL: M21174; AAA30990.1; -.
DR PIR: A40009; A40009.
DR PROSITE: PS01185; CTCK_1; 1.
DR PROSITE: PS01225; CTCK_2; 1.
DR PROSITE: PS01208; WMFC; 1.
DR PFM: PF00007; Cys_knot; 1.
KW Glycoprotein; Repeat.
FT NON_TER 1
FT DOMAIN 1 368 81 AA TANDEM REPEATS.
FT REPEAT <1 44 1 (FRAGMENT).
FT REPEAT 45 125 2.
FT REPEAT 126 206 3.
FT REPEAT 207 287 4.
FT REPEAT 288 368 5.
FT REPEAT 369 391 6 (INCOMPLETE).
FT REPEAT 929 995 WMFC.
FT DOMAIN 1062 1146 CTCK.
```

```

FT DISULFID 1062 1109 BY SIMILARITY.
FT DISULFID 1076 1123 BY SIMILARITY.
FT DISULFID 1085 1139 BY SIMILARITY.
FT DISULFID 1089 1141 BY SIMILARITY.
FT DISULFID 2 1145 BY SIMILARITY.
FT CARBOHYD 46 46
FT CARBOHYD 50 50
FT CARBOHYD 51 51
FT CARBOHYD 57 57
FT CARBOHYD 58 58
FT CARBOHYD 61 61
FT CARBOHYD 66 66
FT CARBOHYD 67 67
FT CARBOHYD 73 73
FT CARBOHYD 74 74
FT CARBOHYD 76 76
FT CARBOHYD 77 77
FT CARBOHYD 81 81
FT CARBOHYD 83 83
FT CARBOHYD 87 87
FT CARBOHYD 91 91
FT CARBOHYD 93 93
FT CARBOHYD 94 94
FT CARBOHYD 96 96
FT CARBOHYD 98 98
FT CARBOHYD 101 101
FT CARBOHYD 103 103
FT CARBOHYD 104 104
FT CARBOHYD 106 106
FT CARBOHYD 107 107
FT CARBOHYD 108 108
FT CARBOHYD 110 110
FT CARBOHYD 114 114
FT CARBOHYD 117 117
FT CARBOHYD 123 123
FT CARBOHYD 124 124
FT CARBOHYD 418 418
FT CARBOHYD 547 547
FT CARBOHYD 917 917
FT CARBOHYD 985 985
FT CARBOHYD 1002 1002
FT CARBOHYD 1068 1068
SO SEQUENCE 1150 AA; 109615 MW; F7C53CE CRC32;

```

```

alignment_scores:
Quality: 104.00 Length: 405
Ratio: 0.654 Gaps: 20
Percent Similarity: 39.259 Percent Identity: 21.975

```

alignment_block:

US-09-323-427-1 x APMU_PIG ..

Align seg 1/1 to: APMU_PIG from: 1 to: 1150

```

23 GTCGTATATACACCGACTGACGACACACAAACAAACAAACAAAC 72
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
769 ValThrValGlyThrThrValAlaIleProGlySerPheAsnThrIysAlaIh 785
73 AACACACACAGCATATATACCCCATCAAGTGAGAGACAGACAGAG 122
|||||:|||||:|||||:|||||:|||||:|||||:|||||
785 rThrProThrGluValAlaIleAlaThrGlyAlaGly..... 797
123 CATCTTAATTTTCTTAAATCGAATTACTTAATCTTGTGAATGATG 172
797 ..... 797
173 ATTCGTATTATGCTTTCTGTACTACACTTAATGCTTATTCGAT 222
|||||:|||||:|||||:|||||:|||||:|||||:|||||
798 .....ThrIleValGlyAlaIleThrSerIleSerIh 807
223 TCCGGTTCGAATGTGTGAA.....GGTGAAGCCAGAAATTG 260
:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

807 rGlyIleSerThrGlyProGluAsnSerThrProGlyThrGluThrG 824
261 AATGTGACCACTTCATATACAAATCATTTTAAATACAGT...AATGCA 307
|||||:|||||:|||||:|||||:|||||:|||||:|||||
824 IySerGlyThrThrSerSerProGlyIleValIySerGluAlaIhThr 840
308 TTCGAAGACATGTTATATGAAAGCTTTATGATCAACAGAGTTGCCG 357
|||||:|||||:|||||:|||||:|||||:|||||:|||||
841 PheIySgly..... 843
358 TAATGATGAAGTGACGTCACATTCGCCGATTTCA.....CTTC 398
844 .....ValGlyThrThrGluAlaGlyIleSerSerGlyAsnSerP 857
399 CATTTGATTCATGCAATGTGCGGTACACAGATCTGCAATCCAGCGGT 448
|||:|||||:|||||:|||||:|||||:|||||:|||||
857 ro...GlySerGlyGlyValThrSerSerGluIuGlyThrSerIhGlu 872
449 ATTTTGTACACAACTGTTGTCATTTGCTTTCATTCATTAATTTGTAC 498
|||||:|||||:|||||:|||||:|||||:|||||:|||||
873 AlaSerGluThrThr..... 878
499 CAAGTTGATCGTCATATCGAGTACATGCTTT...TACATGGAAGCTG 545
|||||:|||||:|||||:|||||:|||||:|||||:|||||
879 .....AlaProArgIleSerAlaThrGlySerThrSerValS 891
546 ATTAACAGTTAGTGCACAGATGAGTATCTGAA...ATCACACTGCT 592
:|||||:|||||:|||||:|||||:|||||:|||||:|||||
891 erIySgluIleIhThrAlaSerProIySValSerSerProGluIhThrAla 907
593 TTYCAACTCA..... 604
908 GlyAlaThrGluAspGlnGluAsnGlyAsnIySThrGlyCysProAlaPr 924
605 .ATTGCCGATGCCAGTATGCCGTTATCAAAATTTGGATGGGACCAA 653
:|||||:|||||:|||||:|||||:|||||:|||||:|||||
924 oleuProProProIyValCysHis.....GlyProI 935
654 CCGGTACACAGTTCATTAATTCATATGTCATGTCACGACGATTCATTA 703
|||||:|||||:|||||:|||||:|||||:|||||:|||||
935 euGlyIuGluIySserProGlyAspValIhThrIhAlaAsnIySHisIyS 951
704 TGGCATGTC...GATTCGAACCGTTGAT.....ACTTTCGCCG 741
|||||:|||||:|||||:|||||:|||||:|||||:|||||
952 CysThrCysThrGluAlaIySThrValAlaSpCysIySProIySgluSPr 968
742 GGTTCATCTCCTGCTTGTGCATGAT.....GGTA 773
:|||||:|||||:|||||:|||||:|||||:|||||:|||||
968 oSerProProThrCysIySThrGlyGluArgIleuIleIySPhelYsAlaA 985
774 ACAGTGATACGTGGAATTCCTAAATGCTGATGATGCTCTTGTATAA 823
|||:|||||:|||||:|||||:|||||:|||||:|||||
985 snAspThrCysCysgluIle.....GlyHisCysgluIySArgIhThr 998
824 TATTTGCTAATTAATTTGGAATATCCACAGATTAATGCTGGCCAGA 873
|||||:|||||:|||||:|||||:|||||:|||||:|||||
999 CysIeuPheAsnAsnThrAspIyr..... 1006
874 AGCTCACGTATACAAATATGCGGATGATCACAGCTTTTCTATCAATGCC 923
1007 .....G 1007
924 AGATCAGTATTAACATTAAAGCAATAGGAATGTGTGACACAA 973
:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1007 IuValIySerSerPheAspAspProAsnAsnProCysValIhThrIySer 1023
974 TGTTCAGAACACACAGATTCGAGCTGTTAA.....AA 1007
|||:|||||:|||||:|||||:|||||:|||||:|||||
1024 Cys...GluAsnThrGlyPheThrAlaValAlaGluAsnIySProIySgl 1039
1008 CAGGTGTGCGCC.....AGCAAAACCTGCTGC. 1035
:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1039 nThrIhTrpCysAlaGluGluAspArgValIyrAspSerIySgIySgIyS 1056

```

```

1036 ..AGCTGGCAACTTCGTTTACTCAAGAAAGACTCGAAGACCGAGAA 1083
      ::::: ::::: ::::: :::::
1056 yRthCysLysSerSerCysLysProSerProValAsnValThrValarg 1072
      ::::: ::::: :::::
1084 TATCATTCATGATGACG 1098
      ::::: :::::
1073 TYRAsnGLYCYSThr 1077

seq_name: SwissProt_38:NU5M_APIILI
seq_documentation_block:
ID NU5M_APIILI STANDARD; PRT; 554 AA.
AC P34855;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3).
GN N5.
OS Apis mellifera ligustica (Common honeybee).
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Prexygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
OC Apoidea; Apidae; Apis.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=THORAX;
RX MEDLINE 93114603.
RA CROZIER R.H., CROZIER Y.C.;
RT "the mitochondrial genome of the honeybee Apis mellifera: complete
sequence and genome organization.";
RL Genetics 133:97-117(1993).
CC -I- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; L06178; AAB96805.1; -
DR PFM; PF00361; oxidored_g1; 1.
DR PFM; PF00662; oxidored_g1_N; 1.
KM Oxidoreductase; NAD: Ubiquinone; Mitochondrion.
SQ SEQUENCE 554 AA; 65590 MW; 9E256E2C CRC32;

alignment_scores:
Quality: 103.00 Length: 554
Ratio: 0.396 Gaps: 34
Percent Similarity: 46.931 Percent Identity: 21.480

alignment_block:
US-09-323-427-1 x NU5M_APIILI ..
Align seg 1/1 to: NU5M_APIILI from: 1 to: 554

465 CTGTGTCATTCGTTTCATCCATATTGTTACCAAGTCATGCTGCA 514
      ::::: ::::: ::::: :::::
11 LeuPheGluPheSerPheLeuMetLeuMetSerLeuTyrLeuLeuTyr 27
      ::::: :::::
515 TATC.....GAGTACATGCTTTTACATGAGAGCTGA 546
      ::::: :::::
27 IleuSlnYsgLuphePhePheGluTrpAsnIleTyrThrPheAsnSerW 44
      ::::: :::::
547 TAAAA..... 551
      :::::
44 eLlyPheAsnPheLeuLeuLeuIleAspTyrLysSerLeuMetPheIle 60
      :::::
552 .....CAGTTAGTCACAGATTGAGCTA 574
      :::::

```

```

61 PheLeuValSerMetIlePheSerMetIleIleIleTyrSerIleSerTyr 77
75 TCTGAATACACAACTGCTTTTCAACTCAAAATGTCGCCAGTATG 624
      ::::: :::::
77 rMetAsp.....LeuSerGluLeuLysMetAspArgPheLeuTyrL 91
      :::::
625 CCGTTATGAAATTTTGGATGTGAGCAACCGTCACACAGTTCAATTWG 674
      :::::
91 eumetIleLeuPhe..... 95
675 CTATCATTTGGTCAGCCAGTTTATCATTAATGACATGCGATTCTCAAC 724
      :::::
96 .....LeuIleSerMetTyrMetLeu.....IleLeuSerPr 106
725 GTGATACTTCTCGCGGTTGTCATTCCTGCTTGCGATG..... 767
      :::::
106 oAsnMetLeuSerIleIleLeuGlyTrpAspIlyLeuIleIleSerT 123
768 .....ATG 770
123 yrcysLeuValIleTyrTyrMetLysMetLysSerPheThrSerGlyMet 139
771 GTAAGGATGATCTGTGGAATTTCAATGCTGATGATGTCCTGTGAT 820
      :::::
140 ValThrIleLeuLeuAsnArg.....LeuGlyAspIleGlyLeuLe 153
821 AAATATTTGCTAAATATAT.....TGAATATACAA... 851
      :::::
153 uLeuLeuMetGlyLeuMetThrTyrTyrGlySerThrPrsLeuSerPheT 170
851 ..... 851
170 yrlYsMetAsnGluPheMetIleTyrIleLeuLeuMetAlaPheThr 186
852 .....CAGATT.....TAATGCTGGCCAGACAGACCTC... 878
187 LysSerAlaGlnIleProPheSerThrTrileuProMetAlaMetMetAl 203
879 ...ACGTATCAAAATATGCGGATGATCAGCTTTTCTATCAATGCA 924
      :::::
203 aprothrProValSerSerLeuValHisSerSerThrLeuValThrAlaG 220
925 GATCAGTATTAACATTAAGAACAAATAGCGAATGTGTCACCACAAT 974
      :::::
220 lYleTyrLeuLeuLeuArg.....TyrValAsnLeuLeuAsnPheAsn 234
975 GTTCAGAACCCAC..... 986
235 TyrLysAsnTyrIleMetLeuIleAlaSerIleThrMetLeuPheAlaG 251
987 .....AAGGATTCGAGAGCTG...TTAAACAGGTGTCGCCAGACANA 1026
      :::::
251 yLeuValAlaAsnPheGluLeuAsnLeuLysLysValAlaIleTyrSerT 268
1027 ACCTGCCTGCGCGCAACTTCGTTTACTCAAGAAAGATGTCAGAGA 1076
      :::::
268 hrLeuSerGlnLeu...GlyPheMetLeuSerMetLeuSerIleGlySer 283
1077 CGGAGAAATATCATGTATGATGACACTGATATCAACACCTTGAATTAC 1126
      :::::
284 ThrGluLeuValPheLeuHisLeuPheIleHisAlaMetPheLysSerIle 300
1127 GATGATTAATCAAGCTTTCGACGTTGATTCATGTCACCGGTCACCTTCTGCA 1176
      :::::
300 uMetPhe.....MetCysValGlySerTyrMetHisTyrMetTyrSerA 315
1177 ACATAATGACAACTGTAATACTGCTGCAG.....TACAAATGGA 1220
      :::::
315 snGlnAspIleArgMetTyrTyrGlyMetTyrTyrIleTyrPrometLys 331
1221 TCTGCATGTCACCAATTTGG...CTTCTCAATGTTATGAGTTAAGCAT 1267
      :::::
332 SerMet.IleLeuIlePheSerIleLeuSerLeuLysGlyPhePheProphe 348

```

```

1268 GCATGATGTCGCGCGCATTAACATTTGCTTAATTTGCTCCAAA 1317
    :::: ||||| |||||
348 euValgLy.....TyrTyr.....SerLys 354
1318 TCAGAGGCGATAAATAATAGTAGATCGAAG.....CATTA 1358
    :::: ||||| |||||
355 Aspleuilellelelumetphepserlymetletypheserme 371
1359 ATAAACTGCCATATA.....TATCGTTTCTTC 1387
    ||| ||| |||||
371 ttleasleuilellelyThrilepethrvalSerTySerphergm 388
1388 TTATCATCCTTTAATAATATTTAGTAACAA.....ATATATAGT 1431
    :::: ||||| |||||
388 etlleleuValleuThrserlypheumetelsvaliletyrSer 404
1432 ATGTAGAA.....ATAATAGTAATAACATAAGTATTTTC 1472
    :::: ||||| |||||
405 Lyglaasplysilemetcysilesermetmetmetletilepese 421
1473 A.....TCAAACTTCTCTATCGCTTTATAGCTTCTGAAAAGCT 1513
    | ||||| ||||| |||||
421 tleuileTySerLysleuilepheasnleuMet.AsnPhe..... 434
1514 TATTCATTTATTCAGTATCTTTTATATGCACTATTTGTAATGTTTCA 1563
    ||| ||||| |||||
435 AsnleuendlylleasnleuMetletyTyr.....LysleuMetVa 449
1564 CATTAGGCCATGATAGTTTCGTTTGTAT.....A 1595
    :::: ||||| |||||
449 lPheLyMetileMetValglyleuileMetglypheasnPheTyLysL 466
1596 TCATCATTTATCAAC.....TTGTCCATTTTATTTAAGCTTTATCAT 1639
    :::: ||||| |||||
466 eulleleuasnlysnlylelyTyRphelymetSer.Pheleuph 482
1640 TTGTGATATATCAACAATTATACCTGTATTCGCCAATTTTATGGCA 1689
    ||||| |||||
482 ewet...AsnleuileTyRlyleTyRlyLysLysLysleileMetMetw 498
1690 TCATTTCCATTTCTGTA.....ACAATTCATTTTTCATATTTGCA 1733
    :::: ||||| |||||
498 etheThrTyRlygluValTyRleleLysSerlelleleleuLeuSer 514
1734 ATTAAT 1739
    |||
515 SerLys 516
seq_name: SwissProt_38:IF2_BACSU
seq_documentation_block:
ID IF2_BACSU STANDARD: PRT: 716 AA.
AC P17889: 031757:
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE TRANSLATION INITIATION FACTOR IF-2.
GN INF2.
OS Bacillus subtilis.
OC Bacteria, Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168:
RX MEDLINE: 90236932.
RA SHAZAND K., TUCKER J., CHIANG R., STANMORE K.,
RA SPERLING-PETERSEN H.U., GRUNBERG-MANAGO M., RABINOWITZ J.C.,
RA LEIGHTON T.;
RA "Isolation and molecular genetic characterization of the Bacillus
RT subtilis gene (infB) encoding protein synthesis initiation factor
RT 2."
RL J. Bacteriol. 172:2675-2687(1990).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-168:
RX MEDLINE: 93259931.
RA SHAZAND K., TUCKER J., GRUNBERG-MANAGO M., RABINOWITZ J.C.,
RA LEIGHTON T.;
RA "Similar organization of the nusa-infB operon in Bacillus subtilis
RT and Escherichia coli."
RL J. Bacteriol. 175:2880-2887(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-168:
RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
CC -! FUNCTION: IF-2, ONE OF THE ESSENTIAL COMPONENTS FOR THE INITIATION
CC OF PROTEIN SYNTHESIS IN VITRO, PROMOTES FORMYLMETHIONYL-TRNA FROM
CC SPONTANEOUS HYDROLYSIS AND PROMOTES ITS BINDING TO THE 30S
CC RIBOSOMAL SUBUNIT. IT IS ALSO INVOLVED IN THE HYDROLYSIS OF GTP
CC DURING THE FORMATION OF THE 70S RIBOSOMAL COMPLEX.
CC -! SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -! ALTERNATIVE PRODUCTS: IT IS PROBABLE THAT BY USING ALTERNATIVE
CC INITIATION CODONS IN THE SAME READING FRAME, THE GENE TRANSLATES
CC INTO TWO ISOZYMES: ALPHA AND BETA.
CC -! SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M34836; AAA22673.1; -
DR EMBL: Z18631; CAA79234.1; -
DR EMBL: Z99112; CAB13536.1; -
DR PIR: A35269; A35269.
DR PIR: B35269; B35269.
DR PIR: S31994; S31994.
DR HSSP: P02990; IETU.
DR SUBTILIST: BG10268; INF2.
DR PROSITE: PS01176; IF2. 1.
DR PFAM: PF00009; GTP_EFTU; 1.
DR Initiation factor; Protein biosynthesis; GTP-binding;
KW Alternative initiation.
FT CHAIN 1 716 INITIATION FACTOR IF-2-ALPHA.
FT CHAIN 94 716 INITIATION FACTOR IF-2-BETA (PROBABLE).
FT DOMAIN 220 368 G-DOMAIN.
FT NP_BIND 226 233 GTP (BY SIMILARITY).
FT NP_BIND 272 276 GTP (BY SIMILARITY).
FT NP_BIND 326 329 GTP (BY SIMILARITY).
FT CONFLICT 56 56 R -> A (IN REF. 3).
SQ SEQUENCE 716 AA; 78621 MW; ACB9B730 CRC32;

```

alignment_scores:

Quality:	102.50	Length:	520
Ratio:	0.418	Gaps:	23
Percent Similarity:	47.115	Percent Identity:	18.077

alignment_block:

US-09-323-427-1 x IF2_BACSU ..

Align seg 1/1 to: IF2_BACSU from: 1 to: 716

```

28 TATACACCGACTGCAGCAGCAACAACAACAACAACAACAACA 77
   :::: ||||| |||||
99 pheasnlysasnllylsasnasnasnlnlylsasnllysaArgAAsnAs 115
78 CAACAGCATATATAC.....CCCATCAAGTGGAGGAGA 112
   ||||| ||||| |||||
115 nasnasnasnlysasnglnhlsGlnGlnLysProVallySProlYsLysG 132

```

```

113 AGACAGAGCAATCTTAGTTTCT..... 138
132 IuleuProgluLysIleThrPheSerGlyThrLeuThrValGlyAlaLeu 148
139 AAAATCGAATTCTAATCTTGAATGATGATTCGT.....C 179
149 AlaGlu.GluLeuGlyLysGluProSerGluLeuIleLysLysLeuMetL 165
180 TTATTCCTTCTGTACTACA..... 199
165 euleuGlyValMetAlaThrIleLeuGlnGluLeuAspLysAspThrIle 181
200 ..CTTATGCAATGCTCTTATTCGATCCGGTGCATGCGTC.... 241
182 GluLeuIleAlaSerGlyTrpGlyValGluThrGluGluValIleValLe 198
242 .....GAAGTGCAGCCAGAA 257
198 uGluGluThrGluLeuGluLysTyrGluGluProAspAsnGluAspL 215
258 TTGAATGGACCAACTTCATTAACATC..... 286
215 euGluIleArgProGlyValThrIleMetGlyHisValAspHisGly 231
287 .....AATTTAATACAGTATGCAATGCAAGGA.. 316
232 LysThrThrLeuLeuAspSerIleArgTyrThrLysValGluGlyG 248
317 .....CATGTTATGTGAAGTCTTATGATCAAG 347
248 uAlaGlyGlyIleThrGlnHisIle.....GlyAlaTyrGlnIleG 262
348 AAGGTGGCGTAATGATGAGGTGACGTCAAGTCCGGAATTCACAT 397
262 Lu.....GluAsnGlyLysIleThrPheLeuAspThr 273
398 CCATTGATGCAATGCTGCCGTACAGATCTGTAATCCACGCG 447
274 Pro.....GlyHisAlaIleAlaPheThrThrMetArgAlaArgL 286
448 TATTTTGTACACACACTGTTCAT.....TCGT 479
286 yAlaGluValThrAspIleThrIleLeuValAlaAlaAspAspGly 303
480 TTCATCATTATTTGTACCAAGTGTGATCGTATCGATGACATGAC 529
303 alMePProGluThrValGluAlaIleAsnHisAlaLysAlaGluVal 319
530 TTTTACATGGAAGCT.....GA 546
320 ProIleIleValAlaValAlaAsnLysIleAspLysGluSerAlaAsnPro 336
547 TAAACAGTTAGTCACAGATTGAGTATCTGAATCAACACTGCTTTC 596
336 PArgValMetGlnGluLeuThrGluTyrGlyLeuValProGluAlaTrp 353
597 AAACCTCAATTGTCCGATGCCAGTATCCGTTAATAATTTTGGATGT 646
353 LysGlyGluThrIlePheValProLeu.SerAlaLeuThrGlyLysGly 369
647 GGACCAACCGGTCAACAGTTCATTTGCTATCATGTCGACCCAGTTTA 696
369 eaSpLLeuValGluMetIleLeuLeuValSerGluValGluGluLeuL 386
697 TCATTAATGAGCATGCGATCTGAACCGTTGATCTTCTGCGGGGTG 746
386 yAlaAsnProAsnArgGlnAlaLysGlyThrValIleGluAlaGluLeu 402
747 TCCAT.....TCCGCTTTTGCATGATGATGATAC.. 775
403 AspLysGlyArgGlySerValAlaThrLeuLeuValGlnThrGlyThr 419
776 .....GGTATACTGTGAAATTCATAATGCTGATGATGATGCTCTG 818

```

```

419 uHisValGlyAspProIleValGlyLysThrPheGlyArgVal..... 434
819 ATAATATTTCTAATAATATTG..... 841
435 ..ArgAlaMetValAsnAspIleGlyArgArgValLysThrAlaGlyPro 450
842 GAATTCACACAGATTTA.....ATGCGTGCCA 870
451 SerThrProValGluIleThrGlyLeuAsnAspValProGlnAlaGlyAs 467
871 AGAAGCTCAGTATCAATAATAGCGGATCGATGACACAGCTTTTCTCAT 920
467 pGlnPheLeuValPheLysAspGluLysThrAlaArgSerValGlyGluA 484
921 GCCAGATCAGTATTACATTAAGAACCAAAATAGCGAATGCTTGACCA 970
484 IaArgAlaSerLysGlnLeuGluGlnAlaArgSerAsp.....LysAla 498
971 CAATGTCAGAACCAACAGATTCGAGCTGTTAAACAGTGCT..... 1015
499 LysLeuSerLeuAspAspLeuPheGluGlnIleLysGlnGlyAspVal 515
1016 .....GCCGCAAGCAAAACCTGCTGAGCTGGC 1043
515 sAspIleAsnLeuIleValLysAlaAspArgGlnGlySerAlaGluAlaL 532
1044 AACTCGTTTACTCAAGAAAGATCTGCAGAACCGGAATATCATTCAT 1093
532 euThrAlaAlaLeuGlnLysIleGluValGluGlyValLysValLysIle 548
1094 GTACGACTGATATGACACACCCCTTGAATTCGATGATCAAGCTTT 1143
549 IleHisThrGlyValGlyAlaIleThrGluSerAsp..... 560
1144 GCCAGTTGATTTACCTCAACCGTGCACTTCTGCACATATGACACAAC 1193
561 .....I 561
1194 TAAATCTGCTCAGTACAAATGGAATCTGCATGTCACCATTTGCTTC 1243
561 IleIleLeuAlaSerAlaSerAlaIleValIle.....GlyPhe 574
1244 TCAATG 1249
575 AsnVal 576
seq_name: SwissProt_38:WUCS_BOVIN
ID WUCS_BOVIN STANDARD; PRT; 563 AA.
AC P98091;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SUBMAXILLARY MUCIN-LIKE PROTEIN.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SUBMAXILLARY GLAND;
RX MEDLINE: 90370871.
RA BHARGAVA A.K., MOIRACH J.T., DAVIDSON E.A., BHAVANANDAN V.P.;
RT Cloning and cDNA sequence of a bovine submaxillary gland mucin-like
RT protein containing two distinct domains.
RL Proc. Natl. Acad. Sci. U.S.A. 87:6798-6802(1990).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: SUBMAXILLARY MUCOSAE.
CC -1- SIMILARITY: TO PORCINE APOMUCIN.
CC -1- SIMILARITY: CONTAINS 1 WFC DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).

```



```

472 LeuHisPheGlyGlnLysAspValLeuAspValTyrGlnThrPheSerPr 488
919 ATGCCGAGATGATATGACATTT.....AAGACCAAAATAGCGAT 959
488 ocSyrLeuSerLeuArgPheAspPheThrLysGlnLysLeuIleThr 504
960 GTGTTCGACCAC..... 971
505 GluPheProGlnSerSerLeuIleAlaGlnLysGluIleAsnLeuGln 521
972 .....AATGTGACAGACCAAGATTCGAGC 999
521 rAsnGlyPhePheGlnThrLeuHisMetArgHisHisThrSerLeuGlu 538
1000 TGTAAACAGGTG..... 1013
538 leuProIleIleLysCysIleLysSerLeuSerThrAspIleIleLeu 554
1014 ..GTGCCGACGCAAAACCTGCTGCAGTCGCGCACTTCGTTTACTCAAG 1060
555 SerIleProLeuLysAsnIle.....ThrTyrValIleSerTh 567
1061 AAAAGTCTGCAGACCGGAGATATCATGATGATAGACAACTGAAATGCA 1110
567 rLysProValProAsnSerLysIle.....TyrAspValSerGlu 581
1111 CACCCCTGAATTAAGGATGATATCAAGCTTTCACAGTTGATTTACGTC 1160
581 aLrPheLeuHisThrSerMetIleIleSerAlaValAsn..... 593
1161 ACCGTGCACCTTGCACAAATATGAGAACCTGTATACTTGGTCGAGTA 1210
594 ....AsnAspCysLys.....ProTyr 599
1211 CAAATGGAATGCAATGTCACCATTTGGCTTCTCAATGTTATGGGTTT 1260
599 rGlnGlySerAlaIleHisGln..... 607
1261 AAGCATTCGATTCGTCGCGTCATTTATACCATTTGTTAATTTTC 1310
608 ..... 614
1311 GTCCAAATCAGAGCGATAAATAATGTTAGATCATCGAAGCATTAAT 1360
615 ThrValProArgArg..... 619
1361 AAAATGCGCATATATA.....TTCGTTCTCTTATCATCCTTTAATA 1404
620 .GlyCysProTyrCysSerSerValValLeuSerTyrAspGlnSerGln 636
1405 ACATAATTTAGSTACAAATATATAGTATGAGAAATATATCTGTAAT 1454
636 Lys.....PheGlnSerMetTyrIleThrAspThrTyrValGln 649
1455 ACAATAGATGATTTTCATCAAACTCTTATGATTCGTTTATAGCTTC 1504
650 .....GluAsnLeuPhe.....ThrGlnHisSerProPhePheLys 662
1505 TGAAGAGCTT...ATTCAATTTACGTAATCTTTTATGACATACATA 1551
662 rGluAsnLeuHisIleHisTyrLeuIleLeuMetAsnGlnTyrValI 679
1552 TAAATGTTTCATCATT...AGGCGATGAATAGTTCGTTGTTATATC 1597
679 leGluValArgGlyAlaTyrArgAlaArgLeuValAsnPheIleVal 695
1598 ATCATATGACATCTGTCATATTTATTAACAGTTTATCATTTGAGAA 1647
696 IleMetVal.....PheIleLeuPheLeuValGlyLeuTyrIle 708
1648 AATGCAAAATTAATACCTGTATTT 1672

```

708 uLeuTyrLysLeuPheValTyrLeu 716

seq_name: SwissProt_38:YAY3_SCHPO

seq_documentation_block:

```

ID YAY3_SCHPO STANDARD; PRY; 649 AA.
AC Q10211;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHEtical 74.5 KD PROTEIN C4H3.03C IN CHROMOSOME I.
GN SPAC4H3.03C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA MURPHY L., HARRIS D., BARRELL B.G., RALANDREAM M.A., WALSH S.V.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z69380; CAA93342.1; -.
DR KW Hypothetical protein; Transmembrane.
FT TRANSMEM 265 285 POTENTIAL.
FT TRANSMEM 564 584 POTENTIAL.
FT TRANSMEM 626 646 POTENTIAL.
SQ SEQUENCE 649 AA; 74488 MW; F04F8763 CRC32;

```

alignment_scores:

```

Quality: 97.50 Length: 403
Ratio: 0.554 Gaps: 23
Percent Similarity: 43.672 Percent Identity: 22.333

```

alignment_block:

US-09-323-427-1 x YAY3_SCHPO ..

Align seg 1/1 to: YAY3_SCHPO from: 1 to: 649

```

239 GTCCGAGTGCAGCGCAATTTGAATGTGGACCACTTCATATAACATCA 288
36 LeuAspGlySerValGlnMetCysTrpPro.....As 47
289 TTTTAATACAGTAATGACATTC.....GAAGAC 317
47 nPheAspSerProSerIlePheHisAlaGlyLeuAspAlaArgAlaGln 64
318 ATGTTATGTGAAGGCTTTTATGATCAAGAAAGGTTGCCGTAATGAGA 367
64 lAsPheSerIleThrProIle...GluGlnThrSerCysLys..... 76
368 GGTGGAGCGTCAGAGTTCGCGCAATTCACCTTCATTTGATTCATGCAAT 417
77 .....GlnMetTyrGlnProSerThrAsnI 85
418 TCGCGCTACAGCATCTGCAATCCAGCTGATATTTTGAACAACAAGTG 467
85 eLeuHisThrLysPheTyrSerGlnArgGlyValLeu.....ArgL 99
468 TTGTCAATTTTCGTTTCAT.....CCATTTATTTGTT 496
99 euleuAspPhePheHisArgProTrpGluAspTyrGlnProLeuTyrPro 115

```

```

497 ACCAAGTTCGTCGA.....TATGAGT 522
    : : |||:
116 TrpLeuIleArgValSerCysIleArgGlyThrSerArgIleGly 132
    : : |||:
523 ACAATGCTTTACATGGAACCTGATAAAGATTAGTGCACAGATTGAGG 572
    : : |||:
132 uGluCysPheProAlaLeuAspTyrAlaArgGlnSerHisGlnThrArgV 149
    : : |||:
573 TATCTCAATACACA...ACTGCTTTCAACCAATGTCGGAGCCA 619
    : : |||:
149 alSerIysIleThrGluAsnTyrTyrGlnAlaGlnPheValProAlaSer 165
    : : |||:
620 GATGCCGTTATGAATTTTGATGGTGCACACCGGTCAA..... 661
    : : |||:
166 GlyAspProLysTyrIleLeuAspCysValProSerGlyAspGlnLeu 182
    : : |||:
662 .....CCAGTCAATTTGCTATCATTTGCTGACG 689
    : : |||:
182 sIleAspLeuGlnLeuIleTyrProAlaGlnHisLeuIleGlnGlyG 199
    : : |||:
690 CAGTTTATCATTAATGACATCGCATTCGAAACCGTTGATACCTTCTGC 739
    : : |||:
199 LyValIleSerTyrLeuGlnLeuGlnGlnGlnGlnIleThrPhe... 214
    : : |||:
740 GCGGTTGTCATTCCTGCTTTGTCGATGATGTAACGGTATACCTGGA 789
    : : |||:
215 .....IlePheArgGlnGlnGlyLeuGlyProAsnValAs 226
    : : |||:
790 AATTCATAATGCTGATGATGCTCTCTGATAATATTCTCAATAATTT 839
    : : |||:
226 pTyrValThrProAsn.....LeuValAspLysL 236
    : : |||:
840 TGGAAATATCCACACAGATTAAATGCTGGCCACAGACTCAGTATACAA 889
    : : |||:
236 euGlnAspSerThrLysArgTyrTrpArgAlaTrpIleGlnGlnCysVal 252
    : : |||:
890 TATGCGATGATCACAAGCTTTTCTATCAA.....TGCAGATCAG 930
    : : |||:
253 TyThrGlyArgTyrArgGlnPheValGlnArgAsnAlaLeuThrLeu 269
    : : |||:
931 TATTAACATTAAAGAACCAATAGCGAATGCTT...CGACCAATGTT 977
    : : |||:
269 sLeuLeuIleTyrGlnProThrGlyAlaValIleAlaSerProThrPheS 286
    : : |||:
978 CAGAACCCACAGAGATTGCGAGCTGTTAAACA..... 1009
    : : |||:
286 erLeuProGluAspLeuGlyGlyValArgAsnTrpAspTyrArgPheThr 302
    : : |||:
1010 .....GGTGGTCCCGCAGCAAAACCTGCTGACGTCGCACTT..... 1048
    : : |||:
303 TrpIleArgAspSerAlaPheThrIleTyrAlaLeuAlaGlnLeuGlyPh 319
    : : |||:
1049 .....CGTTTACTCAAGA 1061
    : : |||:
319 eaArgAlaGlnAlaValGlnTyrMetSerPheIleTyrHisValLeuLysL 336
    : : |||:
1062 AAAGATCTCGCAACCGGAG...AATATCATTGATGTACGACTGATATC 1108
    : : |||:
336 yslYAsnLysAspGlyGlyIleAsnIleVal..... 346
    : : |||:
1109 AACACCCCTTGAATTTAGCGATGATAATCAAGCTTTG...CCAGTTGATT 1155
    : : |||:
347 .....TysSerIleHisGlySerSerGlnAsnLeuGlnGlnValAla 361
    : : |||:
1156 AGGTGAC...CGTGCACTTCTGCAACATATGACAAACCTGTAATCTGC 1202
    : : |||:
361 uThrHisLeuAlaGlyGly.....TyrTyrAsnSerHisProValArgIle 375
    : : |||:
1203 CTGCAATACAAATGCAATCTGCATGTCACCAATTTGCTTCAAGATT 1252
    : : |||:
376 .....Gly.AsnAlaAlaValHisHisLeuGln.LeuAspIleT 388
    : : |||:
1253 ATGGG 1257

```

```

seq_name: SwissProt_38:YCF2_MARPO
388 yIGly 389
|||||
seq_documentation_block:
ID YCF2_MARPO STANDARD: PRT, 2136 AA.
AC P09975;
DT 01-MAR-1989 (Rel. 10, created)
DT 01-MAR-1989 (Rel. 10, last sequence update)
DT 01-JUL-1993 (Rel. 26, last annotation update)
DE HYPOHEMICAL 259 KD PROTEIN (ORF 2136).
GN YCF2.
OS Marchantia polymorpha (Liverwort).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiopsida;
OC Marchantiales; Marchantiaceae; Marchantia.
RN [1]
RP SEQUENCE FROM N.A.
RA OHYAMA K.;
RL Submitted (OCT-1986) to the EMBL/GenBank/DBJ databases.
RN [2]
RP COMPLETE GENOME.
RA OHYAMA K., FUKUZAWA H., KOHCHI T., SHIRAI H., SANO T., SANO S.,
RA UMESONO K., SHIKI Y., TAKEUCHI M., CHANG Z., AOTA S., INOKUCHI H.,
RA OZeki H.;
RT "Chloroplast gene organization deduced from complete sequence of
RT Liverwort Marchantia polymorpha chloroplast DNA.";
RL Nature 322:572-574(1986).
CC -I- SIMILARITY: TO A SIMILAR ORF IN OTHER PLANTS CHLOROPLAST.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X04465; CAA28078.1; -.
DR PIR: A05037; A05037.
DR PIR: S01591; S01591.
DR MEDEL: 5295; MARPO:ycf2.1.
DR PFAM: PF00004; AAA; 1.
KW Chloroplast; Hypothetical protein.
SQ
SEQUENCE 2136 AA; 25991 MW; A2F8B07 CRC32;

alignment_scores:
Quality: 97.00 Length: 662
Ratio: 0.316 Gaps: 33
Percent Similarity: 46.375 Percent Identity: 18.278

alignment_block:
US-09-323-427-1 x YCF2_MARPO ..
Align seg 1/1 to: YCF2_MARPO from: 1 to: 2136
49 AACAAACAA.....CACAAACACAAAC.....AA 77
|||||
709 AsnAsnLysLeuIleThrTrpLysLysIleSerAsnLysLeuValIle 725
|||||
78 CAACACCAATTAATACCCATCAAGTGAAGAGAGACAGCAAGCAATC 127
|||||
725 rAsnSerGluTyrAsnLysIleIleThrAsnLysLysAsnMetLysPhe 742
|||||
128 TTAGTTTCTTAATAATCGAATTTACTAAATCTCTGAATGATGATTGG 177
|||||
742 heSerPheSerLysAsnSerValLeuAspThrPhePheAsnLysLys 758
|||||
178 TCATATGCTTTCTGCTAGTACACTTATTGCATTTGCTTATGATTCCGG 227
|||||
759 SerPheAsnIleIle.....ThrValIlePheAspLysL 770

```

228 TTGACATGCTGCGAAGGTGAGCCGAATTTGAATGTGGACCACTTCA 277
:: :: :: |||||:: ::
770 eulysylsilegluleuasnphleglnlelglulysileleuasnycs 786
278 ATAACATCAATTTTAATACAGTAATGCATTGCAAGACATGTTATGT 327
:: :: :: |||||:: ::
787 pheSerleuphepheaSernerylSasnilellysThrlsillepely 803
328 GAAAGGCTTTATGATCAAGAGGTGGCGTAAATGATGAAGTGGACGTC 377
:: :: :: |||||:: ::
803 sasnseryThrpheileasnleu..... 810
378 AAGTGGCCGAATTTTCATCTTCATTGATTCAATGATTTGGCGCTACA 427
|||:: :: |||||:: ::
811 815
428 CCATCTCTAATCCACGT.....GCTATTTTGTACACACACCTGTGT 471
|||:: :: |||||:: ::
816 pheSerPheasnAsplysglupheasnillephe..... 826
472 CATTCGTTTCATTCATTTTGTGTACCAAGTTGATCGCATATCGAG 521
|||:: :: |||||:: ::
827PheleugluleupheileSerleuleasnAsnAspPheleu 841
522 TACATATGCTTT.....TACATGGAAGCTGATAAACAAGTTAGT 559
:: :: |||:: :: |||||:: ::
841 eArpghpheelyslystYrleuTYrArGlleTYrlySAsplysGl 857
560 GCACATGAGGATATCGAATCACACTGCTTTCAAACTCAATTTGT 609
:: :: :: |||||:: ::
857 uileupheasnProilegluasnArgluleuLeuglnAsnPhpheb 874
610 CCCGATGCCAGTATGCCGTTATGAAATTTGGATGGTGACCAACCGTC 659
:: :: :: |||||:: ::
874 lulystrlySilleleuthrPheilleaspheleuglnAspProgluleu 880
660 AACCACTTCAT.....TTGCTATCATTTGGTCAGCCAGTTATCAATAA 703
|||:: :: |||||:: ::
891 AsnTYrAsnAsnArpheeliePheHlsleuGlulysThrllelySAs 907
704 TGGACAT.....GCCATTTCTGAACCGCTGATACCTTCTGCGC 741
|||:: :: |||||:: ::
907 nAsnAsnleuLeuTYrleuArgluleuLysillePheleuLysAsplySAs 924
742 GGTTCGCAATTCCTGCTTGTGATGATGTAACGGTGAATAGTGGAAA 791
|||:: :: |||||:: ::
924 Tg..... 924
792 TTCTAATGCTGATGATGCTCTTGATTAATATTTGCTAATAATATTG 841
:: :: |||||:: ::
925AsnPhleuleuleasn..... 930
842 GAATATCCACAGATTTAATGCTGGCCAAAGACCTCAGCTATACAAATA 891
|||:: :: |||||:: ::
931 GlullelySer.....Pheilleglulyslys.....AsnAsn 942
892 TGGCGATGCATCACAGCTTTCTATCAATGACGATCAGATTACATTA 941
:: |||:: :: |||||:: ::
942 eupheillelySernleuSer.....AsnVallleuVal 954
942 AAGAACCAA.....ATAGCAATGTGTCGACAC..... 971
|||:: :: |||||:: ::
955 lysAsnSerTYrlySphPheAsnAsnillePheAsnPhelHlsPheleu 971
972 990
971 sglulysglulysAsnilleglulleleuAsnAsnGlnAsnTYrPheb 988
991 ATTGAGAGCTGTAATAACAGGTGTCGCGACAAAACCTGTCGACGCTG 1040
:: :: :: |||||:: ::
988 lulySserleuLeuLyslysthrTYrleuLysasn.....leuAsnleu 1002

1041 CGCAC.....TTGCTTTACTCAAGAAA 1063
:: :: |||:: :: |||||:: ::
1003 AsnAsnSerTYrSerlySpheserTYrlySillePheillePheleu 1019
1064 AGATCTGCAGAACCGGAGATATCATTTGATGACAGACGATGATCAAC 1113
:: :: :: |||||:: ::
1019 uAsnilleuAsnlySAsnAsnTYrlysthrPheGlnTrpIleSer.... 1034
1114 CCTGAATATTAGCATGATTAATCAACCTTGGCAGTGTGATTACGTCACC 1163
1035GluulellePheTYrSer 1040
1164 GTGCACCTTGCAACATAATGACAACCTGTATACTCTGCGATGACA 1213
|||:: :: |||||:: ::
1041LysAsnleuAsnTYrlySilleglulAsnly 1050
1214 AATGGA.....TCGCAATGTCACCATTTGGCTTCTCATATGTTATGG 1257
|||:: :: |||||:: ::
1050 silleglulysAsnAsnTYrCysTYrAsn..... 1059
1258 TTTAACATTCATTCATTCGTCGCCGTCAATTATTACATTTGCTTAAAT 1307
1059 1059
1308 TTCTGCCAATCAGAACGATAAATAATGTTAGATCATCGAACCAAT 1357
:: :: |||:: :: |||||:: ::
1060LysAsnilleSerTYrlySlySlySilleYsthrValAsnPhp 1075
1358 AATAAAGCTGCCATATATATTCGTTCTTCTTATCATCTCTCAATAACT 1407
:: |||:: :: |||||:: ::
1075 eglulys.....AsnAsnleupheGlnThrAsnAsn 1086
1408 AATTAGCTAACCAATATATAGTATGTAGAAATATTATAC.....TGT 1451
:: :: |||:: :: |||||:: ::
1086 exTrpPheThrlleuGlulTrpGlulYrAsnThrlleleu 1102
1452 AATACATAAGTGAATTTTCATCAAA...ACTTCTTCATGCGCTTTAT 1498
:: |||:: :: |||||:: ::
1103 GlnilleleglulnThrPhePheGlnilleThrAspVallleuYrPh 1119
1499 AGCTTGTGAAGCTTATTCATTTATTCAGTAATCTTTATATGCAACT 1547
:: :: |||:: :: |||||:: ::
1119 eLYsLYsLYsLYsilleleuLysAsnleuLysPhePheleuLYsSer 1136
1548 ..ATTGTAATGTTTCATCATTTAGCCATGAATAGTTTC..... 1584
:: :: |||:: :: |||||:: ::
1136 yslsilleSerleuLysThrleu.SerPheHlsAsnPhelysleuYstr 1152
1585GTTTGTATTATCATCATTTATCAACTTG 1612
1152 pAsnleuArpPhePheasnleuGlulleasnTYrlySlySAsnTYrleuVal 1169
1613 TCCTATTATTATCT.....AACAGTTTATCA 1638
:: :: |||:: :: |||||:: ::
1169 snPheleuTrpSerAspPheasnleuileasnAsnycsAsnAsnleuTYr 1185
1639 TTGTGATATATACCAAAATTATACCTTGATGGCCCAATTT..... 1680
:: |||:: :: |||||:: ::
1186 TrpValllePheSerleuValillePheillePheleuTYrGlnlyS 1202
1681TTATGGGATCATTTTC 1696
1202 ePheSerillelleleglYserAspCysPheHlsleuTrpLYsAsnPhp 1219
1697 CTATTCTGTAAACAATTCATTATTGCAATTATGCAATTAATAAGTATT 1746
|||:: :: |||||:: ::
1219 lullelle.....GlnTYrleuThrAspArGserArGserleu 1231
1747 TCATTGTGAAAAAATAAAAAAATAAAAAA 1776
|||:: :: |||||:: ::
1232 TYrPheThrlsleuThrArGArGAsnlyS 1241
seq_name: swissprot_38:GALU_PSEAE

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL: L05499; AAA33163.1; -.
 DR EMBL: L05496; AAA33163.1; JOINED.
 DR EMBL: L05497; AAA33163.1; JOINED.
 DR EMBL: L05498; AAA33163.1; JOINED.
 DR PIR: B42239; B42239.
 DR HSSP: P26769; IAB8.
 DR DICTYDB: DD02024; ACAA.
 DR PROSITE: PS00452; GUANYLATE CYCLASES; 1.
 DR PFAM: PF00211; guanylate_cyc; 2.
 KM Lyase, CAMP synthesis; Transmembrane.
 FT DOMAIN 1 219 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 220 240 POTENTIAL.
 FT TRANSSEM 244 264 POTENTIAL.
 FT TRANSSEM 276 296 POTENTIAL.
 FT TRANSSEM 304 324 POTENTIAL.
 FT TRANSSEM 325 345 POTENTIAL.
 FT TRANSSEM 353 373 POTENTIAL.
 FT DOMAIN 374 962 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 963 979 POTENTIAL.
 FT TRANSSEM 992 1012 POTENTIAL.
 FT TRANSSEM 1018 1038 POTENTIAL.
 FT TRANSSEM 1071 1091 POTENTIAL.
 FT TRANSSEM 1105 1125 POTENTIAL.
 FT TRANSSEM 1378 1398 POTENTIAL.
 FT DOMAIN 1399 1407 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 65 74 POLY-GLY.
 FT DOMAIN 535 576 ASN-RICH.
 FT DOMAIN 753 785 POLY-ASN.
 FT DOMAIN 847 864 POLY-ASN.
 FT DOMAIN 88 113 ASN-RICH.
 FT DOMAIN 532 576 ASN-RICH.
 FT DOMAIN 753 872 ASN-RICH.
 SO SEQUENCE 1407 AA; 159699 MW; F1B7939C CRC32;

alignment_scores: Quality: 96.00 Length: 423
 Ratio: 0.505 Gaps: 20
 Percent Similarity: 44.917 Percent Identity: 19.622

alignment_block:

US-09-323-427-1 x CYAA_DICDI ..

Align seg 1/1 to: CYAA_DICDI from: 1 to: 1407

```

25 GCTCTTAACACCGACTGCAGCAACAACAACAACAACAACA 74
   :: ::::|::|::|::|::|::|::|::|::|::|::|::|
841 lleleuasphrthrvValasnasnasnasnasnasnthrasnms 857
   ::::|::|::|::|::|::|::|::|::|::|::|::|
75 CAAC.....AACAGCATTAATACCCCA 97
   |||
857 nasnasnlysaenglnasnasnlietyrclysnasnasnasnglug 874
   |||
98 TCAACTGAGAGAAGACAGAGACAATCTTAAATTTCTAAATGCA 147
   ::::|::|::|::|::|::|::|::|::|::|::|::|
874 luasphelysileyserserAsnserSer.Phe...Glnilegl 889
   ::::|::|::|::|::|::|::|::|::|::|::|::|
148 ATTACTTAATCTTCTGAATGATGATCGTCTTATGCTTCTGAC 197
   ||::|::|::|::|::|::|::|::|::|::|::|
889 uetserasnlielyslyrsProlyserArg.....Pheileaspa 903
   ||::|::|::|::|::|::|::|::|::|::|::|
198 CACTATTGATGATTGCTATTGATTCGGTACAAATGCTGCAAGST 247
   ::::|::|::|::|::|::|::|::|::|::|::|::|
903 rglValmecclyleleuhsishlsvalysliIleSerAsnsplyl 919
   ::::|::|::|::|::|::|::|::|::|::|::|::|
248 GAGCGAAGAAATTGAATGTGGACCACTTCAATACATC..... 286
   ::::|::|::|::|::|::|::|::|::|::|::|::|

```

```

920 LysglullelelelnlleaspluasphelysvalthrLysleuar 936
287 .....ATTTAATACAGTATGATTCGAG 314
936 glystyrPheTyrrPhePhegluasnasneuthrthrLysPheNLS 953
   ::::|::|::|::|::|::|::|::|::|::|::|::|
315 GACAGCTTTATGTGAAGGCTTTATGATCAGAAAGGTTCGGTATGAT 364
   ::::|::|::|::|::|::|::|::|::|::|::|::|
953 ystyrVallelleasnasnvalvalgluthrLysPhePhe..... 966
   ::::|::|::|::|::|::|::|::|::|::|::|::|
365 GAAGGTGACGTCAAGTTCGGCAATT..... 391
   |||
967 .....LeuVallelleleuNLSleuMetPheTyrlde 979
392 .....TCACTTCATTTGATTCATTCATTCGTTGGCC 422
979 uaspsphisliemetaspserseralProtyrrPheasnserAsnvalileT 996
423 GTACAGATCTGTGAATCCAGCGTATTTTGTAAACACACTGTTC 472
   |||||
996 yrlleu.....Valmetglyllealaphleuvaltyrillegly 1008
473 ATTTGCTTTCATTCATTTATTTGTACCAAGTTGATCGCATATCGAGT 522
   ::::|::|::|::|::|::|::|::|::|::|::|::|
1009 LeuserPheThrArgillePheArgthrProleu.....Valtyrclnll 1023
523 ACAATGCTTTACATGAGAGCGATGAATAACGTTAGTCACAGATTGAG 572
   ::::|::|::|::|::|::|::|::|::|::|::|::|
1023 ealapherlelleleuLycysalaphleglyalCysrthrValleuL 1040
573 TATCTGAATCACA...CTGCTTTCAACCTCAAAATTTGCCCA...TG 616
   ::::|::|::|::|::|::|::|::|::|::|::|::|
1040 eu.lleArgphlelnasproleuAlaArgserleuthrArgValcy 1056
617 CCAGTATGCCGTTATGAATTTGTGATGGTGAGACCAACCGGTCAACCACT 666
   |::|::|::|::|::|::|::|::|::|::|::|
1056 s.....AlathrleuPheTyrlleuasnvalPheNLSerleuAsnphel 1071
667 TCAATTTGCTATCATTTG.....GTACGCCAGTTTATCATTAATGCA 707
   ::::|::|::|::|::|::|::|::|::|::|::|::|
1071 euserValleuPheleuAsnleuPheillePheSerPheheilleleCys 1087
708 CATGCGATTCGAAACCGTTGATTTCTCGCGGTTCCATTCCTCTGC 757
   ::::|::|::|::|::|::|::|::|::|::|::|::|
1088 SerilleuNLSerProthrleuthrAsnNLSleuTyrlthrAsply 1104
758 TTTGTCGATGATGTAAACGCGTACTGTGAATTTCAATGCTGATGG 807
   ::::|::|::|::|::|::|::|::|::|::|::|::|
1104 rilleglyPheValle..... 1109
808 ATGTGCTCTTGAATAATATTGCTTAATATTGGAATTCACACAGAT 857
   |||||
1110 ..ValleuNLSerleuNLSer.....SerSerTyrlglyMet 1121
858 TTAATGCTGCGCAAGACTCAGATATACAAATATGCGAGTCAATG...CA 904
   ::::|::|::|::|::|::|::|::|::|::|::|::|
1122 LysleuAlaMetArglysalatrPvalvalasncyslylileasnphely 1138
905 CAGCTTTTCTATCAATGCCAGATCAAGTATTAACATTAAGAACCAATAG 954
   ::::|::|::|::|::|::|::|::|::|::|::|::|
1138 sThrilleSerValasNLS..... 1144
955 GCAATGTGTGACGACCAATGTTTCAGAACCAACAGAGATTGCGAGCTGTA 1004
   ::::|::|::|::|::|::|::|::|::|::|::|::|
1145 .....GlnLysAsplysPheNLSnPhelLeu 1153
1005 AAACAGGTGTCGCCGACGCAAAACCTGCTGACGTGCGCAACTGCTTTA 1054
   ||::|::|::|::|::|::|::|::|::|::|::|
1154 LysSerillePheProglinseralaleuthrLysleuArg..... 1166
1055 CTCAGAAAGATCTGCAGACCGGAGAAATATCATTTGATGTACAGATCA 1104
   ::::|::|::|::|::|::|::|::|::|::|::|::|
1167 .....Aspmetl 1169

```

```

1105 TATCAACACCCCTTGAATATGCGATGATTAATCAAGCTTTCAGTTGAT 1154
      ||::||| |||  ||::|
1169 leaspthrPro.....AsnIleGluThrLysGlyIleVal 1180
      ||::|
1155 TACGTCACCGTCAC 1169
      |||||
1181 TyrValGlnProHis 1185

```

```

570 AGGTATCTGAATTCACAACTGCTTTTCAAACTCAAAATGTCCGGATGCCA 619
133 luvallsergluilethrthralapheglnthrlnlevalprometpro 149
620 GTATCCGCTTGAATTTTGGATGCGGACCAACCGGTCAACCACTTCA 669
150 ValCysArgTyrGluIleuAspGlyLysProThrGlyInProIleG 166
670 ATTTCATCATGATGTCAGCCAGCTTATCAATTAATGACATGCGATCTG 719
166 nphelaIthrIleGlyInGlnInValTyrHisLysTrpThrCysAspSerG 183
720 AAACCGTTGATCTTCTCGCGCGGTGTCATTCCTGCTTTCATGATGAT 769
183 luthrValaIspThrPheCysAlaIvalAlaHisSerCysPheValaSp 199
770 GGTAAAGCTGATCTGCGAAATTCATTAATGCGATGATGCTGCTTGA 819
200 GlySerGlyAspThrIleGlnIleuAsnGlnGluGlyCysAlaLeuAs 216
820 TAAATATTTCTAATTAATTTGGAATTCACAGATTTAATGGCTGGCC 869
216 pLysTyrLeuLeuAsnAsnLeuGluTyrProThrAspLeuMetAlaGly 233
870 AAGAAGCTCAGTATACAAATATGCGGATGATCAGCTTTCTATCAA 919
233 lnglnAlaHisValTyrLysTyrAlaAspArgSerGlnLeuPheTyrGln 249
920 TGCAGATCGATTTACATTAAGAACAATAGCAATGCTGTTCGACC 969
250 CysGlnIleThrIleThrIleLysGlnProAsnSerGlyCysProArgPr 266
970 ACAATGTTTCAACACACAGAGATTCGAGCTGTTAAACAGCTGGTCC. 1018
266 oThnCysSerGluProGlnGlyPheGlyAlaValArgProGlyLysSer 283
1019 .GCAGCAAACTGCTGCGACGCTGCGCAACTGCTTACTCAAGAAAAGA 1066
283 lvalAlaPolyLysGlnArgArgCysGlnLeuArgLeuIleLysSer 299
1067 TCTGCAGAACCGGAGAAATATCATTTGATGATGAGAACTGATCAACACCT 1116
300 GLyGlyAspLysArgAsnThrLeuAspValArgThrAspSerAlaLe 316
1117 TGAATTAAGCGATGATATCAACGCTTTCGAGTGAATTAACGTCACCGTG 1166
316 uAspIleSerAspArgAspGlnAlaLeuPrometAspLeuArgHisArg 333
1167 CACTTCTGCAACAT...AATGACAACTGTATACTGCTGCAAGTACAA 1213
333 lA.....ArgHisAlaArgGlyGlnGlnValIleLeuSerProAlaAsn 347
1214 AATGAAATGTCATGTCACCATTTGGCTTCAATGTTTATGGTTAAG 1263
348 GluGlyIleCysMetSerProPheGlyPheSerIlePheMetGlyLeuAl 364
1264 CATTGATGATGATGCTGCCGTCATTAATACCATTTGCTTAAATTCGTC 1313
364 aValAlaIleuAlaIleAlaValAlaValAlaSerPheLysLeuArgPr 381
1314 CAATCAGAAAGCA 1327
381 roGlnGlnLysAla 385
seq_name: sp_invertebrate:Q19707
seq_documentation_block:
ID Q19707 PRELIMINARY: PRT: 389 AA.
AC Q19707:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JAN-1989 (TREMBlrel. 09, Last annotation update)
DE F22B5.3 PROTEIN.

```

```

GN F22B5.3.
OC Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidae; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA Sims M.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELL A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., REIKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULLIVAN J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
DR EMBL: Z50044; CAA90355.1;
SQ SEQUENCE 389 AA: 42983 MW: 56895430 CRC32;

alignment_scores:
Quality: 1382.50 Length: 385
Ratio: 4.177 Gaps: 8
Percent Similarity: 85.974 Percent Identity: 69.610

alignment_block:
US-09-323-427-1 x Q19707
Align seg 1/1 to: Q19707 from: 1 to: 389
200 CTATTTGCAATGCTTTATTCGATTCGCGTTGACAAATGCTGCGAAGTGA 249
13 lvalValAlaSerValSerAlaIleProValAspAsnValGluLysG 29
250 GCCAGAAATTTGAATGTGCGACCAACTTCAATACAAATTTAATACAC 299
29 uProGlnuValGluCysGlyProThrSerIleThrValaAsnPheAsnThr 46
300 GTATGCAATTCGGAAGACATGTTTATGTAAGCGTGTAAATGTCAGAA 349
46 rGAsnAlaPheGlnGlyHisValTyrValLysGlyLeuPheAsnGln 62
350 GGTTCGCCGTAATGATGAGGTGAGCGTCAAGTTGCCGAATTTCACTTC 399
63 GluCysArgAsnAspGlnGlyArgGlnValAlaGlyIleLeuLeuPr 79
400 ATTGATTCATGCAATGTTGCGGCTGACAGATCTCGAATCCACGCTGTA 449
79 oPheAspThrCysAsnValAlaArgThrArgSerLeuAsnProLysGly 96
450 TTTTGTACAAACAACTGTTGTCATTTGCTTCATCATTAATTTATTTAC 499
96 alPheValThrThrThrValAlaValaSerPheHisProGlnPheValThr 112
500 AAAGTTGATCGTCATATGAGTACATCAATGCTTTTACATGGAAGCTGATA 549
113 LysValaIspArgAlaIleTyrArgValaGlnCysPheTyrMetGlnAlaAsp 129
550 AACAGTTAGTGCACAGATTGAGTATCTGAATACACAACTGCTTTCAAA 599
129 sThrValSerThrGlnIleGlnValaSerAspLeuThrThrAlaPheGlnT 146
600 CTCGAATTCGCCGATGCGAGATGCGGTATGAATTTTGAATGGTGTGA 649
146 hrcInAlaValaPrometProIleCysLysTyrGlnIleuAsnGlnGly 162

```



```

650 CCAACGGGTCAACAGTTCATATTTGCTATCATTTGCTACGCGAGTTTATCA 699
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
163 ProthrglyuProvalGlnPheAlaThrIleGlyGlnGlnValTyrHi 179
700 TAAATGACATGCGATTCGTGAACCGTTGATACCTTTGCGGGGTGTCC 749
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
179 slystrPhrcysAspSerGluThrValAspThrPheCysAlaValAlH 196
750 ATTCTGCTTTTGCGATGATGATGATGATGATGATGATGATGATGATGAT 799
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
196 lssSerCysThrValAspAspGlyAsnGlyAspThrValGlnIleLeuAsp 212
800 GCTGATGATGATGCTCTGATTAATATTTGCTAAATATTTGGAATATTC 849
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
213 GluAsnGlyCysAlaLeuAspLysPheLeuLeuAsnAsnLeuGluTyrPr 229
850 AACGATTTAATGCTGGCGCAAGAGCTACGATPACAAATATGCGGATC 899
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
229 othrAspLeuMetAlaGlyGlnGlnAlaHisValTyrLysTyrAlaAspA 246
900 GATCAGACCTTTTCTATCATGCGCATGACATPACATPACATPACATPAC 949
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
246 rSerGlnLeuPheTyrGlnCysGlnIleSerTleThrValLysGluPro 262
950 AATAGCGATGTGTTCGACACCAATGTTTCAGAACCAAGATTCGAGAC 999
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
263 AsnGluGlyCysAlaArgProThrCysSerGluProGlnGlyPheGlyAl 279
1000 TGTTAACAGGTGTGCGCGCAGCAAAACCTGCTGACGCTGCCGACATTC 1049
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
279 aValLysGlnAlaAsnGlnThrAlaGln.....PhePheA 291
1050 GTTATCTGACAGAAAGATCTGCAGAA...CCGAGATATATCATGATGA 1096
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
291 rValLeuLysLysArgSerAlaProValMetGluAsnIleLeuAspVal 307
1097 CGAATGATATCAACACCCCTGAAATAGCGATGATTAATCAAGCTTTGCC 1146
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
308 ArgAlaGluLeuThrThrLeuGluValLeuGluGlyAsn....LeuPr 322
1147 AGTTGATTTAGCTCACCGCTGCACCTTCG.....CAACATAATGAC 1187
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
322 oSerLeuThrArgIleAlaIleValAlaLeuValAlaSerArgGluIleGlyG 339
1188 AACCTGATATCTGCTGCGATGACAAATGGAATCTGCATGCGATCACCATT 1237
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
339 lUAsp.....SerPheArgGlnGluLeuCysIleSerSerPhe 351
1238 GCGTTCTCA.....ATGTTTATGGGTTTAAACATTTGCAATTGATTC 1278
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
352 HisIleSerValValThrValPheLeuGlyLeuThrVal.....PheA 366
1279 TGCCGCTCATTTATAC.....ATTTCGTTTAAATTTGCTCCAATC 1319
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
366 lAlaIlePheIleThrTyrMetIleValSerArgMetValProSerA 383
1320 AGAAG 1324
|||||
383 sPlys 384
seq_name: sp_invertebrate:O18479
seq_documentation_block:
ID 018479 PRELIMINARY; PRT; 433 AA.
AC 018479:
DT 01-JAN-1998 (T-EMBLrel. 05, Created)
DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)
DE 01-NOV-1998 (T-EMBLrel. 08, Last annotation update)
DE CUTICLIN-1.
GN MTCUT-1.
OS Meloidogyne artellia.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Diplogasteria; Tylenchida;

```

```

OC Tylenchida; Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97218031.
RA DE GIORGI C., DE LUCA F., DI VITO M., LAMBERTI F.;
RT "Modulation of expression at the level of splicing of cut-1 RNA in the
RT Infective second-stage juvenile of the plant parasitic nematode
RT Meloidogyne artellia.";
RL Mol. Gen. Genet. 253:589-598(1997).
DR EMBL; X96677; CA65452.1; -.
KW Cuticle.
SQ SEQUENCE 433 AA; 47582 MW; FDE26884 CRC32;

alignment_scores:
Quality: 1313.50 Length: 418
Ratio: 4.042 Gaps: 7
Percent Similarity: 77.751 Percent Identity: 63.636

alignment_block:
US-09-323-427-1 x O18479 ..

Align seg 1/1 to: O18479 from: 1 to: 433

170 ATGATTCGCTTATTCCTTTCTGTAACACTTATTCATGCTCT.... 214
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MetArgLysLeuLeuPheAlaIleGlyValPheValAlaLeuAsnAlaI 17
215 .....TATTCGATTCGCGTTGACAAATGCTGCGAAGTGACGCCAG 254
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17 ePheThrValAlaArgAlaIleProValAspAsnGlyValGlyGluProG 34
255 AAATGAAATGAGGACCAACTTCAATACAAATCAATTTTAAATACAGTAT 304
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
34 lIleGluCysGlyProThrSerTleThrValAsnPheAsnThrArgAsn 50
305 GCATTCGAAGACATGTTTATGTAAGGCTTTTATGATCAAGAAGGTTG 354
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 ProPheGluGluGlnHisValTyrValLysGlyLeuPheAspGlnAlaGly 67
355 CCGTATATGTAAGTGCAGCTCAAGTTGCCGAATTTCCATCTCCATTTG 404
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
67 sArgSerAspGlnHisGlyArgGlnValAlaGlyIleGluLeuProPheA 84
405 ATTTCGCAATGTCGCGCTACACGATCTGGAATCCAGCGGTATTTT 454
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
84 sPserCysAsnValAlaArgThr...AspAlaGluProLysGlyValPhe 99
455 GTAACAAACACTGTTGCAATTCGTTTCATCCATATTTGTTACCAAGT 504
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
100 ValSerThrThrValValIleSerPheHisProGlnPheValThrLysVa 116
505 TGATGTCGATATGAGTACATGCTTTTACATGGAAGCTGATAAACAG 554
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
116 lAspArgAlaTyrArgValGlnCysPheTyrMetGlnAlaAspLysThv 133
555 TTAGTGACAGATGAGGTATCTGAAATCACAACGCTTTTCAACCTCA 604
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
133 aIleSerAlaGlnLeuGlnValSerGluIleThrThrGlnPheGlnThrGln 149
605 ATTGTCGCGATGCCAGTATCCGTTATGAAATTTGATGATGAGCAAC 654
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
150 ValValAlaProMetProValCysLysTyrGluIleLeuGluGlyAlaAla 166
655 CGGTACACCACTGTTCAATTTGCTATCATGTCGATGAGCAGTTTATCAT 704
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
166 uGlyGlnProIleGlnPheAlaThrIleAlaGlnGlnValTyrThrSerG 183
705 GGACATGCGATTCGAAACCGTTGATACTTCTGCGCGGTTGCCATTCC 754
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
183 lYThrCysAspSerGluThrIleAspThrPheCysAlaValAlaHisSer 199
755 TGCTTTGTCGATGATGATGATGATGATGATGATGATGATGATGATGAT 804

```

```

11  |||
200 CysValValAspAspGlyAsnGlyAspThrValGlnIleLeuAsnIUG1 216
805 TGGATGTGCTGATTAATATTTGCTAAATAATATGGAATATCAACAG 854
216 ySeRcSAlaLeuAspLysPheLeuLeuAsnIleGluTyrProTha 233
855 ATTTAATGCTGGCCAAAGACTCAGGTATACAATATGCGATCATCA 904
233 spReuThrAlaGlyGlnGlnAlaHisValTylTylStyrAlaAspGse 249
905 CAGCTTTTCTATATCCAGATCATATTTACATTAAGACCAATAG 954
250 GlnLeuPheTylGlnCysGlnIleSerIleThrIleLysGluProHis 266
955 CGAATGTGTCGACCAATGTTCAGACACCAAGATTCGAGCTGTTA 1004
266 rGluCysAlaArgProLysCysAlaGluProSerGlyPheAsnAlaVal 283
1005 AAACAGT.....GGTCCGACGACAAACCTGCTGACGCTCG 1042
283 ySeValGlyAlaGlyGlyAlaGlyAlaGlyAlaProProAlaAla 299
1042 ..... 1042
300 AlaAlaThrProProAlaAlaAlaProProAlaProIleAlaProAl 316
1043 .....CACTTCGTTTACTCA 1058
316 AlleProAlaArgValGlnArgLeuArgThrArgGlnLeuArgLeu 333
1059 AGAAAAGATCTGCAGAACCG...GAGATATCATTTGATGACACTGAT 1105
333 rGlySarGlyArgAlaAlaProGlyGlnGlnIleLeuAspValaArgTh 349
1106 ATCAACACCTTGAATTAGCCATGATATCAAGCTTGGCCATGATTT 1155
350 PheAsnAlaLeuAspIleSerAspLysProPheAsnLeuProAlaHis 366
1156 AGCTCACCGCTGCTGCACATTAAT.....GGACACACCTGTAATAC 1199
366 eArgHisHisHisLeuMetClnAsnLeuSerAlaGluProGlnAla 383
1200 TTGCTGCAGTACAAATGGA.....ATCTGATGTCACCAATTTGGCTTC 1243
383 rGpHeIleTyrGlnGlyGlnGlnAlaLysMetSerPheAlaGala 399
1244 TCAATGTTTATGGG.TTTAAGCATTCATGATGCTGCCCATATATA 1292
400 AsnHisLeuLeuArgTyrGlnIleCysAsnGlyCysSerHisTrpTyr 416
1293 CCAT 1296
416 LHS 417

seq_name: sp_invertebrate:Q9XVM7

seq_documentation_block:
ID Q9XVM7 PRELIMINARY; PRT; 364 AA.
AC Q9XVM7;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE F53F1.1 PROTEIN.
GN F53F1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Pelodetidae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA BURTON J.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

```

```

RT elegans.";
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERS M.,
RA BONFIELD J., BURTON J., CONNELL M., COSEY T., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCT C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans ";
RL Nature 368:32-38(1994).
DR EMBL; Z81088; CAB03124.1; -.
SQ SEQUENCE 364 AA; 40843 MW; 3FEC30BC CRC32;

```

```

alignment_scores:
Quality: 725.50 Length: 271
Ratio: 3.455 Gaps: 1
Percent Similarity: 77.491 Percent Identity: 47.601

```

alignment_block:

US-09-323-427-1 x Q9XVM7

Align seg 1/1 to: Q9XVM7 from: 1 to: 364

```

221 ATTCCGGTGTGACATAGTGTGACAGTGCAGCAAGTGAATGATGAC 270
49 ValProIleGlnAsnSerLeuTyrGlyAspValGlnValGlnCysAsp 65
271 AACTTCATTAACAATCAATTTTAATACACGTAATGATTCGAGACATG 320
65 rArgThrIleSerValGlnIleLysThrClnLysProPheValGlyVal 82
321 TTTATGTGAAGGTCTTTATGATCAAGAAGTTCGCGTATGATGAAGT 370
82 LePheValLysAspPheAlaSerGlnGlnValLysThrSerArgGlyThr 98
371 GGAGCTCAAGTTCGCGAATTTACATTCATTTGATTCGATGATGTCG 420
99 GlyArgLeuSerAlaPheLeuGlnIleGlnIleGlyLeuCysGlyAlaLe 115
421 GCGTACACGATCTCGAATCCAGCTGATATTTTGTACACACACTGTTG 470
115 uArgGlnArgValLeuAsnProLysGlyLeuAlaValaLargThrThrIle 132
471 TCAATTCGTTTCATCCATATTTGTTACCAAGTTCGATGATGATGCA 520
132 TrIleSerPheHisProTyrPheIleThrLysValAspArgThrTyrAsn 148
521 GTACATGCTTTTACATGGAACCTGATAAACAAGTATGACACAGATTGA 570
149 LeuLeuCysLeuTyrArgGlnSerGlnValThrValAlaAsnIleLe 165
571 GGTATCTGAATACACACTGCTTTTCAAACTCAAAATGTCGCCAGTCCG 620
165 rValAspGlnIleSerThrIleSerTyrAsnValaLysnLeuThrMetPro 182
621 TATGCGGTATGAAATTTTGTGATGGGACCAACGCGTCAACCATGTC 670
182 TrCysThrTyrGlnIleLeuSerGlyGlyProPheGlyGlnuProValGlu 198
671 TTTGCTATCATTTGTCAGCCAGTATTAATGAATGACATGCGATTTTCA 720
199 PheGlyLeuIleGlyGlnGlnAlaTyrHisGlnTrpLysCysAspAsnAs 215
721 A.....ACCGTGTGATACCTTTCGCGGCTTGTCC 749

```



```

1 Val11SerPheHisProLeuSerVal1ThrLysValAlaSpArgAlaTyrAr 17
520 AGTACATGCTTTTACATGGAGCTGATATAAACAGTTAGTGCACAGATTG 569
17 gValGInGysPheTyrMetClnuAlaAspLysThrValSerThrlIn11eg 34
570 AGGTATCTGGAATCACAACACTGCTTTTCAACATCAAAATGTCCGATGCCA 619
34 lVal1SerGluIleThrThrlAlaPheGlnThrGln11leVal1PrometPro 50
620 GTTGGCGCTTAATGAATTTTGGATGTGTGGACCAACCGGTCACACAGTCA 669
51 ValCysArgTyrGlu11leuAspGlyGlyProThrGlyInPro11eg1 67
670 ATTGGTATCATGTGTGCAGCCAGTTATCATTAATGACATGCGATTCTG 719
67 nPheAlaThr11eg1GlnProVal1ThrLysTyrThrCysAspSerG 84
720 AAACCGTTGATCTTCTTGGCGGCTTGTCATTCTGCTTTGTGATGAT 769
84 lThrValAlaSpThrPheCysAlaValAlaHisSerCysPheValAlaSp 100
770 GGTAAAGGAGTACTGTGGAATTTCAATGCTGAATGATGATGCTCTTGA 819
101 GlyAsnGlyAspThrValGlu11leuSerAlaAlaSpGlyCysAlaLeuAs 117
820 TAAATATTGCTAAATATTGGAATATCCA 850
117 pLysTyrLeuLeuAsnAsnLeuGluTyrPro 127

```

seq_name: sp_invertebrate:Q21808

```

seq_documentation_block:
ID Q21808 PRELIMINARY; PRT; 290 AA.
AC Q21808;
DT 01-NOV-1996 (Tremblrel. 01. Created)
DT 01-NOV-1996 (Tremblrel. 01. Last sequence update)
DT 01-JAN-1999 (Tremblrel. 09. Last annotation update)
DE R07E3.3 PROTEIN.
GN R07E3.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA COTTAGE A.;
RL Submitted (May-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARDON N., PERCY C., RIFEEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTSON J.,
RA THIERRY-MEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WIKITSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
DR EMBL; Z49207; CAAB9068.1;
SEQUENCE 290 AA; 33215 MW; 441587FE CRC32;

```

alignment_scores:
Quality: 654.50 Length: 283
Ratio: 3.306 Gaps: 8
Percent Similarity: 69.965 Percent Identity: 45.230
alignment_block:

us-09-323-427-1 x Q21808
Align seg 1/1 to: Q21808 from: 1 to: 290

```

437 AATCCACAGTGTATTTTGTGAACACAACTGTTGTCATTTGCTTATCC 486
4 SerProArgGlyLeuPheLeuSerThrAsnValAlaValAlaPheAsnPr 20
487 ATATTATTGACCAAAAGTTCATGATGATGATGATGATGATGATGATGAT 536
20 ogLupPheLeuThrLysAsnAspArgValPheLysValGlnCysPheTyr 37
537 TGGAGCTGATAAACAAGTATGACAGATTGAGTATGAGTATGAGTATGAC 586
37 etGluMetGluArgGly11eg1LysVal11eg1In11SerMetProPro 53
587 ACTGCTTTTCAACATCAATTCCTCCGATGCCGATGCGGTTATGAAAT 636
54 ProThrMetHisSerLysGlnLeuAsnMetProValCysLysTyrGlu 70
637 TTTGATGATGACACCAACCGGTCACAGTTCATTAATGCTATGATGATG 686
70 lLeuAspGlySerProThrGlyProProVal1ThrPheAlaThrValGly 87
687 AGCCAGTTATCATTAATGACATGCGATTCTGAACCGTTGATATCTTC 736
87 lMetValTyrHisLysTyrThrCysAspThrGlnHisGluAsnThrPhe 103
737 TGGCGCGTGTCCATTCCTGCTTTGTGATGATGATGATGATGATGATGAT 786
104 CysMetLeuValHisSerCysPheValAlaSpGlyLysGlnArgVal 120
787 GGAATTTCTAAATGCTGATGATGCTCTTGTGAATATTGCTTAATAA 836
120 lGlnLeuLeuAsnAspLysGlyCysAlaLeuAspLysTyrLeuLeuTh 137
837 ATTTGGAATATCCACAGATTTAATGCTGCGCCACAGATCAGTATAC 886
137 snLeuGluTyrProThrAspLeuMetAlaGlyArgGluAlaHisValTyr 153
887 AATATGCGGATGATGACAGCTTTTCTATCAATGCGACAGTCACTTATC 936
154 LysTyrAlaAspArgAspAsnMetTyrPheAspCysGln11Ser11leTh 170
937 CATTAAAGAACCAATATCCGAA...TGTGTCGACCCACATATGTTCAAG 983
170 rValLysGluProGlyLeuAspTyrCysAspValProSerCysProAsp 187
984 CA..... 985
187 roProArgArgArgSerAsnThrLeuProAlaProAspAsn11le 203
986 .....CAAGATTGCGAGCTGTTAAAC 1008
204 ThrAla11leAla11leHis11leGluTyrGlu.AspSerGlu11le11Ser 220
1009 AG...GTGGTGGCGAGCAAAACCTGCTCAGCTG.....CAG 1043
220 spTyr11le11leProAsnAspPhe11le11leSerLeuAsn11leGluArg 236
1044 AACTTCGTTTACTCAAGAAAGATCTGCAGAACCGAGATATCATTTGAT 1093
237 AsnPhe.....AspMetArg11leSer..... 243
1094 GTACGACTGATATCACACCTTGAATTAATGAGATATCAACGCTTT 1143
244 ....GluLeuCysMetThrAla11leGlyThrThrLeuLeuValPheLeu 259
1144 GCGAGTTGATTACGTC.....ACCGTGCACCTTTCG 1175
259 snAlaPheLeuPhe11le11leSerLeuValSer11leValHisValCys 274

```

seq_name: sp_invertebrate:Q21540

```

seq_documentation_block:
ID   Q21540      PRELIMINARY;          PRT:   550 AA.
AC   Q21540;
DT   01-NOV-1996 (TReMBLrel. 01, Created)
DT   01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE   M142.2 PROTEIN.
GN   M142.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Pelodierinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MCMURRAY A.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KESHAM J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULLSTON J.,
RA   THIERRY-MEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
RT   *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.
RT   Nature 368:32-38(1994).
RL   EMBL; 273428; CAA97806.1; -.
DR   PFM; PF00092; Cwa; 1.
DR   PRINTS; PR00453; VMEADOMAIN.
SO   SEQUENCE 550 AA; 61445 MW; A055FB45 CRC32;

alignment_scores:
Quality: 605.00      Length: 338
Ratio: 2.713         Gaps: 8
Percent Similarity: 65.976   Percent Identity: 40.533

alignment_block:
US-09-323-427-1 x Q21540 ..

Align seg 1/1 to: Q21540 from: 1 to: 550

137 CTAAATGCAATTACTTAATCTCTGAATGATGATTCGCTTATGCG 186
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
197 VALArgValIphethrValIgluserPheglInLeuAspArgAlaLeuAl 213
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
187 TTTCTGTACTACACTTATGCACTGTGCTTATTCGATTCGCGTTGACAATG 226
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
213 a.....AspserLeuthrIpsrMetCysLysTfhGlup 225
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
237 GTGTGCAAGGTGAGCCAGAAATGATGTGACCACTTCATAACAATC 286
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
225 heArgproglYhrprogluilelleIecysglYproAspArgIleGlyVal 241
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
287 AATTTTAATACACGTAATGCAATTCGAGACATGTTATGTGAAGCTCT 336
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
242 LysAlaserThrIysglInProPheglInLysValIphethrAlaMetAspH 258
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
337 TTATGATCAAGAAGCTTGCGGT.....AATGATGAAGGTGGAC 374
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
258 sTyrHIsAspGluGluCysArgAlaGlyProgluLysPheProAspSer 275
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
375 GTCAAGTCCCGGAATTCATTCATTCATTCATTCATTCATTCATTCATTC 424
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
275 rgsrIle...GlyLeuThrValProPheSerAlaCysAsnValIhIsArg 290
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
425 ACAAGATCTCTGAATCCACGTGGTATTTTGTAAACAACAACGTGTGCAT 474

```

```

291 TYrArgSerLeuAsnProLysGlyIlePheValGluValSerIleValIph 307
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
475 TTTGTCATTCATCCATTAATTGTTATCCAAAGTGTATGCGATATGAGTGC 524
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
307 ekeTrPheHIsSerLeuPheMetThrIysTyrAspGlnThrValIleValG 324
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
525 AATGCTTTTACATGAGTGAAGCTGATTAACAGTTAGTGTGACAGATTTGAG 574
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
324 IncysPheTyrMetGluAlaAspLysHisValIthrValIProleuSerVal 340
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
575 TCTGAATGACAACTGCTTTTCAAACTCAAAATGTCGCCAGCTATG 624
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
341 SerMetIleThrThrValPheArgGluGlnIleTyrGlnMetProGlnCy 357
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
625 CCGTTTGAATTTTGGATGCTGGACCAACCGGTCAACAGTTCATTTG 674
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
357 salATyrrThrLeuArgLysGlyAlaProAspGlyProIleValIArgPhe 374
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
675 CTATCATTTGTCAGCCAGTTTATCATTAATGACATGC.....GAT 715
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
374 IeThrLeuGlyGluSerValTyrHisArgTrpGluCysIleGlyValGlu 390
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
716 TCTGAACCGTTGATCTTCTGCGCGTTGCCATTCCTGCTTGTGCA 765
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
391 GlyAlaAspLysAspPhePheglMetLeuValHIsSerCysTyrValAs 407
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
766 TGAATGTAACGGATACGTCGTGAATTCGAATTCGATGATGTCGTC 815
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
407 pAsnGlyTyrGlyAspArgValAlaAspIleLeuAspSerAsnGlyCysGly 424
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
816 TTGATTAATATTTTGTAAATTAATTTGAATATCCACAGATTTA...ATG 862
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
424 euAspAlaValIleuLeuSerThrProAspPyrAspPheSerIleuArgLeu 440
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
863 GCTGGCCAAAGAACCTCAGTATACAAATATGCGGATGCATCAGCTTTT 912
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
441 AlaThrLysProTyrHisValIphethrValIAspArgProValLeuG 457
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
913 CTATCAATGCCAGATCAGTATTCACATTAA...GAACCAATATGAGAT 959
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
457 nPheGlnCysGlnIleThrLeuCysLeuLysTyrAspGlyGlyCysGluG 474
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
960 GTGTTCGACCAACAATGTTTCAGAACCAAGATTCGAGCTGTAAACA 1009
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
474 LyIleThrPro.....ProGlnAsnCyLysLysLeuProGly 486
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
1010 GGTGTGCCCGCAGCAAAACCTGTCAGCTGCGCAACTTCGTTACTCAA 1059
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
487 GluAspGlyHIsHIsHIsHIsHIsHIsProGluLysArgArgLysLeuVa 503
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
1060 GAAGAAGATCTCGCAACCGGAGAAATTCATTTGATGATGACAACTGAT...A 1106
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
503 IArgArgLeuAlaAspGlyValGlyThrIleAspValIphethrAspSerV 520
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
1107 TCAACAACCTTGAA 1120
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
520 alThrValLeuGlu 524
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::

seq_name: sp_invertebrate:Q19053
seq_documentation_block:
ID   Q19053      PRELIMINARY;          PRT:   484 AA.
AC   Q19053;
DT   01-NOV-1996 (TReMBLrel. 01, Created)
DT   01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT   01-JAN-1999 (TReMBLrel. 09, Last annotation update)
DE   E04D5.3 PROTEIN.
GN   E04D5.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Pelodierinae; Caenorhabditis.
RN   [1]

```



```

14 ThrSerLysAlaIleValHisIleIlePhePheTPTyLeuCysAsnAr 30
197 ...ACACTTATTCATGCTTATTCATGATTCGGTTCGAATGATGTCG 242
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
30 glIeProIeuPheCysLysTyRHisIaGluGlnIleAspAsnGlyLeuG 47
243 AAGGTGAGCCAGAAATTTGAATGTGAGCAACCTTCATAATCAATCAATTT 292
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
47 lngIyIuProIeuIleArGysGlySerGIuSerIeuSerIleAsnPh 63
293 AATACACGTAATGCATTCGAGACATGTTTATGCAAGGCTTTATGA 342
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
64 LysThngInglYAlaPheGlnGlyHisValTyValLysGlyHisTyse 80
343 TCAGAGAGGTGCCGTAATGATGAAGTGCAGTTCAGTTCGCGCAATTT 392
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
80 rIeLysThsCysArgThrAspAlaThrIeuGluSerGIuValAsnIleuT 97
393 CACTTCATTTGATTCATCAATGTTGCGCGTACACAGATCTGAAATCA 442
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
97 hVAlSerTySerIaLysAspValIleArGlnArGSerSerAsnPro 113
443 CGTGGATTTTGTAAACAACAGTGTTCATCTGCTTCATCCATCATTT 492
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
114 LysGlyIleMeMetThrAlaThrIleIleIleSerPheHisProMetPh 130
493 TGTTCACCAAGTTCATCGCATATGCAATGATACATGCTTTACATGAAG 542
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
130 eIleThrLysIleAspLysSerTyRtyValGlnCysPheTyAlaGlnA 147
543 CTGATAAAACTAGTGCACAGATTGAGTATCT ..... 577
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
147 lAgInLysThrValThngInGlnIleuAsnValAspIleAlaLysGln 163
578 .....GA 579
164 GluLysLysIlePheValIleValGlyAspGlnGlyGlyThrValse 180
580 AATTCACAACTGCTTTTCAAACTCAAT ..... 607
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
180 rHsThrThrGlyAspGlnLysLysIleuHisLysLeuAsnAspProSerT 197
608 .....GTCCGATGCCAGTATGCCGTTATGAA 634
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
197 hngIuGlnArgIleSerTyRAsnValProIeuProAspCysLysTyRArg 213
635 AATTTCGATGGTGGACCAACGCGTCAACAGTTCAATTTGCTATCATTTG 684
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
214 ValIeuThrGluSerLysThr...GluGluValAlaIaPheAlaThrValGI 229
685 TCAGCCAGTTTATCATTAATGACATGCGATTCT .....GAATCCG 725
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
229 yGlnIleValTyRHisGlnIuTrpSerCysGlnAlaProGlyGlnAsnGIaT 246
726 TTGATTACTTTCGCGGCTTCATTCCTGCTTTGTCATGATGTAAC 775
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
246 hTSePProPheCysValIthrValHisSerCysAsnValLysAspGluTrh 262
776 GGTGATACGTGGAATTTCTAAATGCTGATGATGATGCTGCTTGATGAATA 825
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
263 GlyLysGlnValGlnIlePheAspGlnAsnGlyCysAlaValAspLysTy 279
826 TTTTGCTAAATTAATTTGGAATATCCACAGATTTAATGCTGCGCAAGG 875
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
279 rIeuIleAsnAsnLeuGlnIuTrpSerSerAspLeuThrGlyGlnLeuS 296
876 CTCACGTATACAAATATGCGGATCGA 901
296 eXIuValCysSerTrpThrValArg 304
seq_name: sp_invertebrate:Q22164
seq_documentation_block:

```

```

ID Q22164 PRELIMINARY; PRT; 384 AA.
AC Q22164;
DT 01-NOV-1996 (Tremblrel, 01, Created)
DT 01-NOV-1996 (Tremblrel, 01, Last sequence update)
DT 01-JAN-1999 (Tremblrel, 09, Last annotation update)
DE T04F8.4 PROTEIN.
GN T04F8.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA LENNARD N.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SOLSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN S., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROUT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z66565; CA91480.1; -.
SQ SEQUENCE 384 AA; 43964 MW; 9EFEFF44 CRC32;

```

```

alignment_scores:
    Quality: 546.50      Length: 394
    Ratio: 2.240        Gaps: 9
    Percent Similarity: 61.929    Percent Identity: 31.218

```

```

alignment_block:
US-09-323-427-1 x Q22164 ..

```

```

Align seg 1/1 to: Q22164 from: 1 to: 384

```

```

179 CTTATGCTTTCGTACTACATTTATGCTGCTTATTCGATCCGCT 228
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10 LeuValGlnPheThrTyRPhaValLeuGlyPheSerAlaIaIleGlnS 26
229 TGACAAATGCTGCGAAGTGAGCCAGAAATTTGAATGTGACCAACTTCA 278
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
26 pAspAsnGlnLeuIleGlyGlnProGluIleGlnCysAsnAlaAspThrI 43
279 TAACATCAATTTTATACACGTAATGCATTCGAGACAGACATGTTTGG 328
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
43 LeaspMetGlnPheArgThrArgLysGlnPheAsnGlyLysValTyVal 59
329 AAAGTCTTTTANGATCAAGAAGGTTGCCGT .....AATGA 363
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 LysGlnSerTyRAsnArgProGluCysArgValAspTySerThrLysAs 76
364 TGAAGTGGAGCGTCAAGTTCGCGGAATTTTCACTTCATTTGATTCATGA 413
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
76 pGlnPheGlyArgProValGlyLysLeuAsnHisGlyAlaCysA 93
414 ATGTTGCGGTACACGATCTCTGAATCCACGCGTATTTTGAACAACA 463
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
93 smetAspArgGlnArgMetIleAlaProGluIleYMetMetPheSerThr 109
464 ACTGTGTCATTTGCTTTCATCATTTATTTGATCAAGATTGATGTCGCG 513
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
110 ValLeuIleIleSerPheHisProIeuPheLeuThrArgMetAspLysAl 126
514 ATATGAGTACAAATGCTTTTACATGGAAGCTGATTAACAGTTATGTGAC 563

```



```

945 AACCAAT.....AGCAATGTTGACACATGT..... 976
    : : : : : : : : : : : : : : : : : :
278 lmetgylmetcysglngluvalthrprobasncysgylvalyls 294
    : : : : : : : : : : : : : : : : : :
977 .....TCAACACACAGAGCTGGAGCTGTTAAACAGAGTGGTCCGC 1020
    ||||| : : : : : |||
295 leulessergluaserglyaspGly..... 303
1021 AGCAAAACCTGCTGACAGCTGGCAACTGCTTTACACAGAAAGATCTG 1070
    : : : : : : : : : : : : : : : : : :
304 .....AsnArgThrL 307
1071 CAGAACCGAGAAATATCATTTGATGTACAGACTGATATCAACACCTTGAA 1120
    : : : : : : : : : : : : : : : : : :
307 ysArggluAlaaspArgserAspTyrGluLeuLeaspAlaIatThrSerGlu 323
1121 ATTAGCATGATATATCAAGCTTTGCCAGTTGATTACGTACCGCTGACT 1170
    : : : : : : : : : : : : : : : : : :
324 Leu.....LeuValLeuAspProAlaaspArgGlyLe 334
1171 TCTGCAACATATGACACACCTGTAATATCTGCTGACAGTACAAATGGAA 1220
    |||||
334 uLeu..... 335
1221 TCTGCATGTCACCATTTGCTCTCAATGTTATG.....GGTTAAGC 1264
    : : : : : ||||| : : : : : |||
336 ..AlaProserProPheCysValProArgLeuLeuProValLeuPro 351
1265 ATTCGATGATTTGCTGCCATTTACCATTTTCGTTAAATTTGCTCC 1314
    : : : : : ||||| : : : : : : : : : :
352 LeuLeuLeuThrIleValSerLeuThrValVal.SerThrAlaLeuV 368
1315 AAATCAGAACG.....CATAAATATATGTTAGATCATCAGAACG 1355
    ||||| : : : : : : : : : : : : : : : :
368 allleatgargluAsnHisLysLysGluLeuAspIleMetGlnSer 383
seq_name: sp_invertebrate:093967
seq_documentation_block:
ID 093967 PRELIMINARY: PRT: 440 AA.
AC 093967: 094405:
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)
DE ZK265.8 PROTEIN.
GN ZK265.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida;
OC Rhabdilitina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA DOBSON R.;
RA Submitted (Oct-1996) to the EMBL/GenBank/DBJ databases.
RM [2]
RP SEQUENCE FROM N.A.
RP MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BOWFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS J., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN T., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THERRER-MEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RM [3]
RP SEQUENCE FROM N.A.
RA LENNARD N.;
RA Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.

```

```

DR EMBL: 281143; CAB03520.1; -.
DR EMBL: 275713; CAB03520.1; JOINED.
DR EMBL: 275713; CAB00057.1; -.
DR EMBL: 281143; CAB00057.1; JOINED.
SQ SEQUENCE 440 AA: 49299 MM; 3F2BBBA2 CRC32:

```

```

alignment_scores:
    Quality: 477.50      Length: 320
    Ratio: 2.296         Gaps: 9
    Percent Similarity: 65.000      Percent Identity: 33.438

```

```

alignment_block:
US-09-323-427-1 x 093967

```

```

Align seg 1/1 to: 093967 from: 1 to: 440

```

```

173 ATTCGCTTATGCTTTCTGTAACACTTATGATTTGATTTGATTCGAT 222
    : : : : : : : : : : : : : : : : : :
4 LeuLysIleLeuIleIleAlaIatThrGlyTTPArgValAlaAsnAlaI 20
223 TCCGGTTGACATGCTGTCGACAGTGCAGAAATTTGATGTGGACAA 272
    : : : : : : : : : : : : : : : : : :
20 eSerIleAspAsnGluIleIleGlyGluProAspIleGlyCysLeuGluA 37
273 CTTCATTAACATCAATTTTATATACAGTATGATTCGACAGACATGTT 322
    : : : : : ||||| : : : : : ||| : : : : :
37 spcIuIleArgIleIleTyrValLysThrArgLysIlePheAlaGlyArgIle 53
323 TATGTGAAGCTCTTTATGATCAAGAGCTTCCGTTAATGATGATGAGT 372
    : : : : : ||||| : : : : : : : : : :
54 TyrAlaLysGlyArgAlaGluLeuGluAspCysTyrLysAspPheGlu 70
373 AGCTCAAGTTGCC.....GGAATTCACCTTCATTTGATTCATGCA 413
    : : : : : : : : : : : : : : : : : :
70 yAsnGlnLysThrArgLysProHisPheAspLeuGlnPheGlyAlaCysG 87
414 ATGTGGCGCTACACGATCTGTATCACTGATGTTATTTTGTACAAACA 463
    : : : : : ||||| : : : : : ||||| : : : : :
87 LysMetLysSerLeuArgSerValAspProArgLysMetYrGlyLe 103
464 ACTGTGTCACTTTCGTTTCACTCATTTATTTGACCAAGTGTGCTGC 513
    : : : : : ||||| : : : : : ||||| : : : : :
104 ThrValValValSerPheHisProLeuPheIleThrLysValAspIleAl 120
514 ATTCGATGATGCTTTTATACATGAGCTGATTAACAGTATGAGCGAC 563
    : : : : : ||||| : : : : : ||||| : : : : :
120 atYrHisValLysCysPhePheGluGluAlaAsnLysGlyLeuThrAlaG 137
564 AGATTGAGTATGTGAATCACAACACTGTTTCAAACTCAAAATTTGCCG 613
    : : : : : ||||| : : : : : ||||| : : : : :
137 LuLeuGlyValSerMetIleProThrThrGluLeuAlaIatArgHisGly 153
614 ATGCCAGTATGCCGTTATGAAT.....TTGCA 642
    : : : : : ||||| : : : : : ||||| : : : : :
154 IleProGlyCysThrYrSerIleHisArgSerThrIleAspGluLeuAs 170
643 TGTGTGA...CCAACGGTCAACAGTCAATTTGATTCATTCATGTCGAC 689
    : : : : : ||||| : : : : : ||||| : : : : :
170 pAlaGlyArgProAlaGlyAsnValIleGlnPheAlaValAlaGlyGluA 187
690 CAGTTTATATTAATGACATGCGATTCGAAACCGTTGACTACTTCTGC 739
    : : : : : ||||| : : : : : ||||| : : : : :
187 rGlyAlaLeuHisGlnTrpHisCysAsnAspGln.....MetYrGly 200
740 GCGGTGTGCATTCCTGCTTGTGTCATGATGATGATGATGATGATGATG 789
    : : : : : ||||| : : : : : ||||| : : : : :
201 ValLeuIleAsnAsnCysTyrValThrAspArgPheGlyLysLysAlaAs 217
790 AATTCATAATGCTGATGATGCTGCTGATTAATTTGCTAAATAAT 839
    : : : : : ||||| : : : : : ||||| : : : : :
217 pValIleAspAspLysGlyCysProIleAspProIleLeuIleThrGlyI 234

```

```

840 TGGAAATTCACAGATTATG...GCTGGCCAGAGACCTCAGTATAC 886
      :::::|||||:::|||||:::|||||:::|||||:::
234 lAArgTysSerSerAspLeuGlnArgAlaTyrAlaGluSerSerValPhe 250
      :::::|||||:::|||||:::|||||:::|||||:::
887 AATATGCGGATGCATCAGACGCTTTCTTCATGCACATCAGTATAC 936
      ||:::|||||:::|||||:::|||||:::|||||:::
251 LysPheAlaAspLysProGlyValAlaTyrPhePheCysGlnValGlnMetCys 267
      :::::|||||:::|||||:::|||||:::|||||:::
937 CATTTAAGAACCAATATAGCAATGT.....GTTCGACGACCAATGT. 976
      :::::|||||:::|||||:::|||||:::|||||:::
267 smetLysLys...HisGlyMetCysAspLysPheTyrPheProPheProSerCysG 283
      :::::|||||:::|||||:::|||||:::|||||:::
977 .....TCAGAACCAAGATTCGAGCTGCTTAAACAGGTGCTGCC 1018
      :::::|||||:::|||||:::|||||:::|||||:::
283 LysMetSerArgValIleSerValGlyGlyLysAspAsnGlyGlyPhe 299
      :::::|||||:::|||||:::|||||:::|||||:::
1019 GCAGCAAAACCTGCTGACGCTGGCAGCTTCGTTACTCATCAAGAAATAC 1068
      :::::|||||:::|||||:::|||||:::|||||:::
300 GluGluGluGluLysAlaProSerSerArgArg.....LysThrThr 313
      :::::|||||:::|||||:::|||||:::|||||:::
1069 TGCAGAACCG 1078
      :::::|||||
313 rProLysPro 316

```

```

seq_name: sp_invertebrate:Q22680

seq_documentation_block:
ID Q22680 PRELIMINARY; PRT; 609 AA.
AC Q22680:
DT 01-NOV-1996 (TRENBLREL 01, Created)
DT 01-NOV-1996 (TRENBLREL 01, Last sequence update)
DT 01-NOV-1999 (TRENBLREL 12, Last annotation update)
DE T22C8.7 PROTEIN.
GN T22C8.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP THOMAS K.;
RA SEQUENCE FROM N.A.
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BOWFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMILDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
DR EMBL; Z49071; CAAB8879.1;
DR HSSP; P03000; ITRF.
SO SEQUENCE 609 AA; 68882 MW; 0BA2EE0C CRC32;

```

```

alignment_scores:
  Quality: 453.50      Length: 287
  Ratio: 2.399        Gaps: 11
  Percent Similarity: 65.854  Percent Identity: 37.282

```

```

alignment_block:
US-09-323-427-1 x Q22680

```

```

Align seg 1/1 to: Q22680 from: 1 to: 609

```

```

169 GATGATTCGCTTATTCGCTTCTGTAACACTTATTCGATGCTTATTT 218

```

```

|||||:::|||||:::|||||:::|||||:::|||||:::
20 AspAsnSerAsnLysTyrAsnPro...TyrTyrThrHis.....Met 31
      :::::|||||:::|||||:::|||||:::|||||:::
219 CGATTCGCGGTGATGCATGCTGCTGACAGGTGAGCCGCAAAATGTAATGTA 268
      :::::|||||:::|||||:::|||||:::|||||:::
31 tGlu.Pro.....ProLysLeuGluCysGly 39
      :::::|||||:::|||||:::|||||:::|||||:::
269 CCACATTCATATACATCAATTTTAAATACAGCTAATGCATTCGACAGACA 318
      :::::|||||:::|||||:::|||||:::|||||:::
40 SerGluGlyIleArgLeuHisIleAsnProThrGlyThrPheGlyLys 56
      :::::|||||:::|||||:::|||||:::|||||:::
319 TGTATTGTGAAGAGCTTTATGATCAAGAGGTGCCGTAATGATGAAG 368
      ||:::|||||:::|||||:::|||||:::|||||:::
56 sValTyrValArgGlyPhePheProGlnThrValCysHisIleAsnTyrC 73
      :::::|||||:::|||||:::|||||:::|||||:::
369 GTGACAGCT.....CAAGTGGCCGAATTCACCTTCATTT...GAT 406
      |||
73 ystThrArgLeuThrAsnArgProIleValMetAspLeuProPheArgGly 89
      :::::|||||:::|||||:::|||||:::|||||:::
407 TCATGCATGTGTGCGGCTACACGATCTGTGAATCCAGTGATTTTGT 456
      |||||:::|||||:::|||||:::|||||:::
90 ProCysAsnValArgArgArgAsnValAlaProProSerIleSerTyr 106
      :::::|||||:::|||||:::|||||:::|||||:::
457 AACACAACTGTTGCAATTCGTTTCATCCATATTGTTACCAAGTTG 506
      :::::|||||:::|||||:::|||||:::|||||:::
106 rAspValThrValIleIleGlnHisHisProLeuPheValThrSerPheA 123
      :::::|||||:::|||||:::|||||:::|||||:::
507 ATGTCGATATGATGATACAAATGCTTTACATGGAAGCTGTAATAACAGTT 556
      ||:::|||||:::|||||:::|||||:::|||||:::
123 sPlysAlaTyrArgLeuAsnCysIleTyrArgGlnGlnGlnSerThrLeu 139
      :::::|||||:::|||||:::|||||:::|||||:::
557 AGTCACAGATTTAGGTATCTGAATC...ACAAGCTTTTCAAACTCA 603
      :::::|||||:::|||||:::|||||:::|||||:::
140 GlnGlnArgIleAsnValSerAspIleProSerThrAlaLeuGlnSerLys 156
      :::::|||||:::|||||:::|||||:::|||||:::
604 AATTGTCGCCGATGCCGATATGCAATTTGGATGGTGACCA 653
      :::::|||||:::|||||:::|||||:::|||||:::
156 sAsnAlaProLys.....CysArgTyrAspValLeuSerGlySerLeuA 171
      :::::|||||:::|||||:::|||||:::|||||:::
654 CCGGTCACACAGTTCAATTTGATATGCTGTCGACGACGATTTTCATAA 703
      :::::|||||:::|||||:::|||||:::|||||:::
171 smGlyProValValArgPheAlaAsnValGlyAspValValAlaHisLys 187
      :::::|||||:::|||||:::|||||:::|||||:::
704 TGGACATGCGATTCGAACCGTTGATCTTCTGCGCGGTGTCATTC 753
      |||||:::|||||:::|||||:::|||||:::
188 TrpThrCysAspSer.....AspArgPheGlyPheValValHisSe 201
      :::::|||||:::|||||:::|||||:::|||||:::
754 CTGCTTTGCGATGATGATGTAACGCTGATCTGTGAAATTTAAATGCTG 803
      |||:::|||||:::|||||:::|||||:::|||||:::
201 rCysValValArgAspGluSerGlyLysAspPheGlnPheIleAspGluA 218
      :::::|||||:::|||||:::|||||:::|||||:::
804 ATGCATGCTGCTGCTGATTAATTTGCTAATTTGGAATTCGACACA 853
      |||||:::|||||:::|||||:::|||||:::|||||:::
218 rGlyLysValThrAspPheSerLeuPheProGluValSerTyrSerAsp 234
      :::::|||||:::|||||:::|||||:::|||||:::
854 GATTATATGCTGCGCCAGAAAGCT...CACGTAACAAATATGCGGATCG 900
      |||||:::|||||:::|||||:::|||||:::|||||:::
235 AspLeuLysSerAlaPheThrAlaValArgAlaPheArgTyrThrAspG 251
      :::::|||||:::|||||:::|||||:::|||||:::
901 ATCAGAGCTTTTATCATGATGCGATCAGATATTACCATTAAGAACCAC 950
      :::::|||||:::|||||:::|||||:::|||||:::
251 nValMetValHisPheSerCysGlnIleThrCysGlnLysGlnGluA 268
      :::::|||||:::|||||:::|||||:::|||||:::
951 ATAGC...GAATGTGTTGACGACCAATGTTCAACAGACAGAGATTCGGA 997
      ||:::|||||:::|||||:::|||||:::|||||:::
268 smGlyCysGluGlyIleSerProIleCysArgProMetAspLeuGly 284
      :::::|||||:::|||||:::|||||:::|||||:::
998 GCTGTTAA 1006
      :::::|||||
285 ProIleLys 287

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 16, 2000, 01:49:51 / Search time 4841.33 Seconds
(without alignments)
-1115.753 Million cell updates/sec

Title: US-09-323-427-1

Perfect score: 1779
Sequence: 1 ggttaattaccacaagtty.....aaaaaaaaaaaaaaaaaaaaa 1779

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_dal:*
2: gb_dal:*
3: gb_dal:*
4: gb_dal:*
5: gb_dal:*
6: gb_dal:*
7: gb_dal:*
8: gb_dal:*
9: gb_dal:*
10: gb_dal:*
11: gb_dal:*
12: gb_dal:*
13: gb_dal:*
14: gb_dal:*
15: gb_dal:*
16: gb_dal:*
17: gb_dal:*
18: gb_dal:*
19: gb_dal:*
20: gb_dal:*
21: gb_dal:*
22: gb_dal:*
23: gb_dal:*
24: gb_dal:*
25: gb_dal:*
26: gb_dal:*
27: gb_dal:*
28: gb_dal:*
29: gb_dal:*
30: gb_dal:*
31: gb_dal:*
32: gb_dal:*
33: gb_dal:*
34: gb_dal:*
35: gb_dal:*
36: gb_dal:*
37: gb_dal:*
38: gb_dal:*
39: gb_dal:*
40: gb_dal:*
41: gb_dal:*
42: gb_dal:*
43: gb_dal:*
44: gb_dal:*

45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_htg4:*
50: gb_pl3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	339.2	19.1	25823	34 CEF22B5	250044 Caenorhabd
2	323.8	18.2	34831	34 CEF47G2	249125 Caenorhabd
3	315	17.7	2310	34 CELCUT1	M55997 C.elegans c
4	263.4	14.8	724	35 AF125580	AF125580 Wucherer
5	249	14.0	5792	34 ALU073005	U73005 Ascaris lum
6	217.2	12.2	2584	34 MAMRCUT1	X96677 M.artificia
7	180.6	10.7	637	34 BMA012617	AJ012617 Brugia ma
8	156.6	8.8	358	34 BPA012618	AJ012618 Brugia pa
9	107.4	6.0	39478	34 CEF53F1	Z81088 Caenorhabd
10	83.8	4.7	31536	34 CEF04D5	Z66496 Caenorhabd
11	74.8	4.2	32412	34 CEF53B6	Z81086 Caenorhabd
12	71.6	4.0	110000	32 CEF11B2_0	Z98857 Caenorhabd
13	68.8	3.9	152878	34 CEF18D10A	AL034393 Caenorhab
14	68.8	3.9	247332	32 CEF18D10	AL008871 Caenorhab
15	67.8	3.8	73229	34 CEF11B2D	AL132885 Caenorhab
16	67	3.8	39736	34 CER07E3	Z49207 Caenorhabd
17	67	3.8	150724	10 AP000550	AP000550 Homo sapi
18	67	3.8	164399	34 PFMA13P6	Z98551 Plasmodium
19	67	3.8	180894	40 AC008018	AC008018 Homo sapi
20	66.8	3.8	12029	35 AE001381	AE001381 Plasmodiu
21	66	3.7	114790	41 AC011430	AC011430 Homo sapi
22	65.8	3.7	151338	32 HSKB152G3	AL132982 Homo sapi
23	65.2	3.7	35077	34 CEF38H8A	AL021483 Caenorhab
24	65.2	3.7	110000	41 AC005140_2	Continuation (3 of
25	65	3.7	2654	34 DDIDG17A	M18106 Dictyostell
26	65	3.7	256172	33 AC005139	AC005139 Plasmodiu
27	64.6	3.6	27694	7 HASMT	D31785 Hansenula w
28	64.2	3.6	110000	32 CEF39A1_3	Continuation (4 of
29	64	3.6	665	9 HUMAPOBXA	J05157 Human apoli
30	64	3.6	683	15 M23044	M23044 Figure 6. D
31	64	3.6	802	10 S61712	S61712 Apo(B) (HVE
32	64	3.6	940	9 HUMAPOBFL	J02775 Human restr
33	64	3.6	1066	9 HSAPOB3	X04682 Human hypov
34	64	3.6	14059	35 AE001390	AE001390 Plasmodiu
35	64	3.6	283925	41 AC010728	AC010728 Homo sapi
36	63.8	3.6	163678	41 AC010178	AC010178 Homo sapi
37	63.8	3.6	195539	41 AC005506	AC005506 Plasmodiu
38	63.4	3.6	211190	44 AC016716	AC016716 Homo sapi
39	62.8	3.5	106650	40 AC007708	AC007708 Homo sapi
40	62.6	3.5	36306	34 CEM142	Z73428 Caenorhabd
41	62.6	3.5	176552	40 AC004617	AC004617 Homo sapi
42	62.4	3.5	683	15 M23045	M23045 Figure 6. D
43	62.4	3.5	2215	35 S39048	S39048 knob assocl
44	62.4	3.5	288703	41 AC006278	AC006278 Plasmodiu
45	62.2	3.5	36069	34 CEFK265	Z81143 Caenorhabd

ALIGNMENTS

RESULT 1
LOCUS CEF22B5/25823 bp DNA INV 02-SEP-1999
DEFINITION Caenorhabditis elegans cosmid F22B5, complete sequence.
ACCESSION Z50044
VERSION Z50044.1 GI:899234
KEYWORDS HTG; Cuticulin; Elongation factor; GTP-binding ADP-ribosylation

SOURCE factor: phenylalanyl-tRNA synthetase: RNA binding.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditidae; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 25823)
AUTHORS Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M., Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J., Coulson,A., Craxton,M., Dear,S., Du,Z., Dublin,R., Favello,A., Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M., Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N., Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B., O'Callaghan,M., Parsons,J., Percy,C., Ritken,L., Koopra,A., Saunders,D., Showkneen,R., Smaildon,N., Smith,A., Sonhammer,E., Staden,R., Sulston,J., Thierly-Mieg,J., Thomas,K., Vaudin,M., Vaughan,K., Waterston,R., Watson,A., Weinstock,L., Wilkison-Sproat,J. and Mohlman,P.
TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2 (bases 1 to 25823)
AUTHORS Sims,M.
JOURNAL Direct Submission
COMMENT Submitted (13-JUL-1995) Louis, MO 63110, USA. E-mail: joesanger.ac.uk or rwenematode.wustl.edu
Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information.
For a graphical representation of this sequence and its analysis see:
http://webace.sanger.ac.uk/cgi-bin/display?db=wormace&class=Sequence&object=F22B5
Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone F22B5. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between submissions.
The true left end of clone F22B5 is at 24607 in sequence 266522.
The true right end of clone F22B5 is at 16002 in sequence 266523.
The true left end of clone M05D6 is at 25723 in this sequence. The true right end of clone F14E5 is at 4610 in this sequence. The start of this sequence (1..101) overlaps with the end of sequence 266522.
The end of this sequence (25723..25823) overlaps with the start of sequence 266523.
FEATURES
source Location/Qualifiers
1..25823
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="III"
/clone="F22B5"
411..1471
/gene="F22B5.2"
join(411..462,584..663,745..1127,1216..1471)
/gene="F22B5.2"
/note="similar to RNA binding protein: cDNA EST yk39048.5 comes from this gene; cDNA EST yk499d6.3 comes from this gene"
/codon_start=1
/protein_id="CA90354.1"
/db_xref="GI:3876226"
/db_xref="SWISS-PROT:Q19706"

gene
CDS
/translation="MAPAPEVVSMAEVEQDNADHIOGADGTETETAFTEDVGRMKVVTOFKYINKRPKVVADRRKKVKEGSGCAPQAVATTVAAVEVMQFTRNAGDILVDYEDKQFATKTSREHCRKCGNDHSTCKPYKVMYQIDDEADADKDFEKDMAGMRPGROIDRRSPDNCRTNLIPOENNEDELDLPFGKIGIVITIFIAARKVITSLPRGFATVFSRDDAARAIAEIENDIRMYHMYLAKEMTRPSN"
1870..2668
/gene="F22B5.1"
join(1870..2158,2209..2387,2534..2668)
/gene="F22B5.1"
/note="similar to GTP-binding ADP-ribosylation factor"
/codon_start=1
/protein_id="CA90353.1"
/db_xref="GI:3876225"
/db_xref="SWISS-PROT:Q19705"
/translation="MGFLKILRKQRAERERIRILILDNAGKTTLKKFLDEPTDI EPTGPDITVHFKFDPLNLDVGCKSLRKYKWFSESTPAITLVVDSRRELLCCSEELKILGEERLGLASLIYANKSDIPGADVSTIAVIONMSVYISEKKCFVLVDHSIKSHMKTFSCALSGDRLYQAMTLCDDVGSRLILD"
complement(2833..4393)
/gene="F22B5.3"
complement(join(2833..3042,3088..3700,3823..4048,4273..4393))
/gene="F22B5.3"
/note="similar to cuticulin: cDNA EST EMBL:T01970 comes from this gene; cDNA EST CEBSX90RB comes from this gene; cDNA EST EMBL:T00412 comes from this gene; cDNA EST yk386c7.3 comes from this gene; cDNA EST yk386c7.5 comes from this gene; cDNA EST yk504b12.3 comes from this gene; cDNA EST yk543g2.3 comes from this gene; cDNA EST yk652b2.3 comes from this gene"
/codon_start=1
/protein_id="CA90355.1"
/db_xref="GI:3876227"
/db_xref="SPTREMBL:Q19707"
/translation="MARYSLGLGLIVASYSALPVDNNGEPEVEGPTSTVFNTRNAPEGHYVYKGLPDDOECNDGCGROVGLIELPPTCNVARTSRSLNPGCVYTTVVSFHPQVTKVDATYRVOCYMADRTVSQIIEVSDLTTRFQVQVPMPLCKEILNKGPIGEVQFATIGQOVYHKWTCSEVVDVFCVAVHSCYVDDGSDTVQIILDENGCLDKPLINLNEPTDLMAGOEAHVYVYADRSOLFQCOQISTYKKEPNEBCARPTCEPQFGAVKQNOTQOFPFRLKRSAPVMENTLDVRAETLLEVGMLPSLTQOQALVNSREIGDSFROELCISSFHSIVVYVFLGIVTVAIFITMYIYSRMVWPSDKMOSAC"
9691..10234
/gene="F22B5.4"
join(9691..9890,9940..10234)
/gene="F22B5.4"
/note="cDNA EST yk304c8.3 comes from this gene; cDNA EST yk304c8.5 comes from this gene; cDNA EST yk594h3.3 comes from this gene"
/codon_start=1
/protein_id="CA90356.1"
/db_xref="GI:3876228"
/db_xref="SPTREMBL:Q19708"
/translation="MSLSKAPRKTRFPEVLSRNASTSNVSPARIOLKPAEACHPOYSRMSRDPREFYKVAIOKGDPIYOFVLRIGAVYVYFLVLTAMFVLPFCSASVMSGKAETWIDRSNKAPWMDERLRLDITYMKPTVAFDLDGRTKRCELMEDLODBMLAARKRGTR"
complement(10849..12154)
/gene="F22B5.5"
complement(join(10849..10984,11033..11343,11397..11521,11578..11832,11917..12154))
/gene="F22B5.5"
/codon_start=1
/protein_id="CA90357.1"
/db_xref="GI:3876229"
/db_xref="SPTREMBL:Q19709"
/translation="MNCCKOSASKVFELVGFKNKASSRCSSSEKILPKDKLYKNPSEQIKQFQKLVDENDIKLRTMKMDQVLTLCKRAKMSIALCTYNLNSNWSITVSDSTVLSGTQSTIYKSLNRSIPDQFTGLSQDPNTYNLPLHLMHEMATDFKSVKNMSGGCVDSKYAAVWARKVLHTKVGSPTYVYSHKPKDEVPFLEFDDVYVFSFVPEHVKYDNIIFPYERREYTLAKFNGVGVAAVXKGEVIGLVESSGNCIGPIICDNKNAOAIPOSITIODLPKDKDLYVNCSPKFDSDATWIRPFLRRSHEMTPFAHINFRVYPDGLNFSKVFYSSNPNAPC"
complement(12240..14032)

```

CDS
    /gene="F22B5.6"
    complement(join(12240..12394,12467..12541,12691..12784,
    13262..13397,13465..13524,13876..14032))
    /gene="F22B5.6"
    /codon_start=1
    /protein_id="CAA90358.1"
    /db_xref="GI:3876230"
    /db_xref="SPTREMBL:O19710"
    /translation="MKNYGTFAYVNSCCSTWISOKRPSYLEALILPTPESNWL
    LAGYHIHESNNIISMNLMSSRGKIYENATHTGRLVTTAFILCLISLIVDL
    FKTHFTFRVYFVNIAVACGLFVTGLLGHIEVPAKMSITYSQGLDGIQYLLTS
    NSTGEYRKLESAGCELTASSEFKNVFFDLLIQAVFIPHLILTAESWTFSGLYRPRW
    HEDFE"
    15384..21812
    /gene="F22B5.7"
    join(15384..15575,15622..16479,16527..17105,18608..18694,
    18794..18946,18996..19449,19497..19843,20042..20554,
    20672..21600,21677..21812)
    /gene="F22B5.7"
    /note="limited similarity with the elongation factor EF-3
    (Swiss Prot accession number P25997): cDNA EST EMBL:U12588
    comes from this gene; cDNA EST yk349b7.5 comes from this
    gene; cDNA EST yk420b2.5 comes from this gene; cDNA EST
    EMBL:M89075 comes from this gene; cDNA EST EMBL:D36040
    comes from this gene; cDNA EST yk421d3.3 comes from this
    gene; cDNA EST yk421d3.5 comes from this gene; cDNA EST
    yk396b3.3 comes from this gene; cDNA EST yk396b3.5 comes
    from this gene; cDNA EST yk359g5.3 comes from this gene;
    cDNA EST yk359g5.5 comes from this gene; cDNA EST
    yk351a5.3 comes from this gene; cDNA EST yk351a5.5 comes
    from this gene; cDNA EST yk355a11.3 comes from this gene;
    cDNA EST yk355a11.5 comes from this gene; cDNA EST
    yk348b2.3 comes from this gene; cDNA EST yk348b2.5 comes
    from this gene; cDNA EST yk297a7.3 comes from this gene;
    cDNA EST yk297a7.5 comes from this gene; cDNA EST
    yk251h3.3 comes from this gene; cDNA EST yk251h3.5 comes
    from this gene; cDNA EST yk200h11.3 comes from this gene"
    /codon_start=1
    /protein_id="CAA90359.1"
    /db_xref="GI:3876232"
    /db_xref="SPTREMBL:O61442"
    /translation="MSNMDYLDEVDILPKLPNFDLRESKWOERKEALEALLKVL
    DNERLSTKASYAELIGHLOMVLAKDANINCOALAKIGFAGLRKFSFGLPLP
    VIEKMKKPKMLREPLDVCNSNGRMTGEOLELALAKPNOICQALFEVA
    ROLDLVPAKQKPGFIKAVYPVFGKLTGDDOVRASIOGLDAVONIIIDCKVKNIL
    GDASDEPKMKRTIGEYAEKSTASFAEQANAPVAPTSSTPSASGSPSGRTAK
    VSSGAPVAPADPMDFLDARDVLSKMPDGDITNESKWKERKEALEGLQITTAPE
    LDRKANVALVERLQVLEKDNINVAALANICITGLANLRTKRFQPFASVPIIIE
    KFEKKPTLRLDPLVACIDAVVATTNLEAVEIYLAALGKPNPSIKTQTDLFDRCEFK

```

Query Match 19.1%; Score 339.2; DB 34; Length 25823;

Best Local Similarity 63.3%; Pred. No. 5.2e-49; Indels 134; Gaps 2;

Matches 633; Conservative 0; Mismatches 233; Indels 134; Gaps 2;

```

Qy 282 caatcaatttaataacgtaatgcatcgtgaagagacatgtaatgtaaaagtcttcatg 341
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4051 CAGTAAACCTTCAACACAGTAAATGCTTCGAGGACACAGCTTACGTTAAGGACTTTTCG 3992

Qy 342 atcaagaaggttcgcgttaatgtaaggtgagctcaagtgcggaatttcaactcat 401
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3991 ATCAACAAGATGTAGAATATGATCAAGAGAGGCTGCTCAAGTGTGCAATCAACTTCAT 3932

Qy 402 ttgattatgaatggttcgcgttaacgattctgtaatccacggtgattatgaacaa 461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3931 TCGATTACTTGCACAGTGTGCTGACGAGATCTGCAATCCAAAGGAGTCTTCGTTACCA 3872

Qy 462 caactgttcaattcgtttcatcatattgttaccaaagtgtatc----- 510
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3871 CAACTGTGTCGTCTTTCATCCCAATTTGTCACCAAGTGTGATCGGTAAAGACTATG 3812

Qy 510 ----- 510

Db 3811 AAACCTGTGTAATTTTAAAGAAATGCTTCTTGAAGTTTAGAATACCGAAATATATAAGC 3752

```

```

Qy 510 -----gtgcatactg 519
Db 3751 TTGATATCACTAAGACGTTCTCGTTAAAAAATGCAAAATGAATTTTCCAGTGCCTACCG 3692

Qy 520 agtaaatgctttacatggaagcttgataaacaagttagtgacagaatttgagatctga 579
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3691 TGTTCATATGCTTTACATGGAAGCCGACAGACAGTGTCTACCCAAATGAAGTCTCGA 3632

Qy 580 aatcaacatgcttttcaactcaaatgtgtccgagtgtccagatgtgcgttgaattt 639
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3631 TCTTACACACCCCTTCCCAACCCAGGTTGTTCCAAATGCCAATCTGTAAATCGAGATTCT 3572

Qy 640 ggaatgttgaccacacggttcacacgaatttcaattgtctatcatcttgtagcgaatttca 699
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3571 TAAGGAGAGACCAACTGGAGACAGGTTCAATTCGCAACTTGTGGACACAGTCTATCA 3512

Qy 700 taatgacatgacgattctgaaacggttgatacttctgcgcggtttgcattctcgtc 759
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3511 CAATGTGACCTGTGATTTCTGAACCGTGTGATACCTTTTGCGCTGTCTGCTTACCTTGAC 3452

Qy 760 tgtcgaatgatacgtgtgatactgtggaattctaaatgtcgtatgatagtgtctctga 819
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3451 TGTGTGATGATGGAATGTGTGATACCGTGCACAAATTTTGGATGAGAACGGTGTCTTGA 3392

Qy 820 taatattgttaataatttggaatatccaaagatttaattgctgtgccaagaagctca 879
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3391 TAAGTTCTTGTCTCAACATTTTGAATATTCACATGATTTGATGGCTGGACAGAGGCTCA 3332

Qy 880 cgtatacaatatacgcgaatcgcatacagctttctatcaatgacagatcagatataccat 939
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3331 CGTGTCAAAATATACGCTGATCTGTCTCAACCTTTTCTATCAATGCCAGATTTTCATTATCGT 3272

Qy 940 taaagaaccaaatagcgaatgtgttcgaccaaagtgtcagaaccacaaggaattcgaagc 999
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3271 TAAAGAACCAATATGAGAGTGTCTGCGCCCAACGTCCTGACGACAGAGATTGGAGC 3212

Qy 1000 tgttaaaacagtggtgtgcgcgacgaacacgtgcgcagctgcgaactctgttaactca 1059
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3211 AGTTAA-----GCAAAGCTAACCAACTGCTTCATGTTCTTCAGAGTTTGAAGAA 3164

Qy 1060 gaaaagatctgcgaacccggaatatcatgtatgatacgaactgatacaaacaccttga 1119
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3163 GAGATCCGCTCCAGTAATGAGAAATATCTTGATGTCCGCTGAGCTCACCAACTTTCGA 3104

Qy 1120 aattaagatgataatacgaacttgcgaagttgattacgt 1159
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3103 AGTTTGGAAAGAAATGTAACTGTCAAAATGTGTTCAT 3064

RESULT 2
CEC47G2 34831 bp DNA INV 02-SEP-1999
LOCUS Caenorhabditis elegans cosmid C47G2, complete sequence.
DEFINITION 249125
ACCESSION 249125.1 GI:790365
VERSION 1
KEYWORDS HMG; Cuticulin; Fork head.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 34831)
AUTHORS Wilson,R., Almscough,R., Anderson,K., Baynes,C., Betks,M.,
Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
Lattelle,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
O'Callaghan,M., Parsons,J., Percy,C., Riffen,L., Roopra,A.,
Saunders,D., Showkneen,R., Smaildon,N., Smith,A., Sonhammer,E.,
Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
Vaughan,R., Waterston,R., Watson,A., Weinstock,L.,
Wilkinson-Sproat,J. and Wohldman,P.

```

TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans

JOURNAL Nature 368 (6466), 32-38 (1994)

MEDLINE 94150718

REFERENCE 2 (bases 1 to 34831)

AUTHORS Palmer,S.

TITLE Direct Submission

JOURNAL Submitted (27-APR-1995) Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rwenematode.wustl.edu

COMMENT Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information.

For a graphical representation of this sequence and its analysis see: -

http://webcbe.sanger.ac.uk/cgi-bin/display?db=wormaceclass=sequence&object=C47G2

Current sequencing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones.

Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone C47G2 is at 1 in this sequence. The true right end of clone C47G2 is at 6388 in sequence 249912.

The true left end of clone T24F1 is at 34731 in this sequence. The true right end of clone T05B9 is at 12750 in this sequence. The start of this sequence (1..101) overlaps with the end of sequence 249129.

The end of this sequence (34731..34831) overlaps with the start of sequence 249912.

FEATURES

source Location/Qualifiers

1..34831

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="II"

/clone="C47G2"

2562..4878

/gene="cut-1"

join(2562..2679,2992..3217,3898..4636,4687..4878)

/gene="cut-1"

/note="Similar to cuticulin: cDNA EST EMBL:C10444 comes from this gene; cDNA EST EMBL:C12078 comes from this gene; cDNA EST yk262c12.3 comes from this gene; cDNA EST yk262c12.5 comes from this gene; cDNA EST yk265h3.5 comes from this gene; cDNA EST yk278g8.3 comes from this gene; cDNA EST yk278g8.5 comes from this gene; cDNA EST yk439e8.3 comes from this gene; cDNA EST yk439e8.5 comes from this gene; cDNA EST yk380a11.3 comes from this gene; cDNA EST yk380a11.5 comes from this gene"

/codon_start=1

/protein_id="CA88934.1"

/db_xref="GI:3875030"

/db_xref="SWISS-PROT:Q03755"

/translation="MTWKPIICLAALVLSAIPVDNNVEGEPEVEECGPNSTIVNNTINPEGVHYVKGIDQAGCRSDGCGROVAGIELPDCNTARTSLNPKGVFTTVYISHPQVTVKDRAYRIQCFMESDKVSTQIEVSDLTFAQOVVMPKGVFTTVIGGPSQPIQPAITIGQVYHKMTCDSETDTCRAVHSGTVDDGNDTQVILNEBCALDKFLNLLEIFPDLMAGQEAHYKYADRSQLEQYQISITIKDPESCARPTSPQGFKAQAQAGAGHAAAPQAGVEEQAAPVAAPVAAAAPAPVAPVATLQRLTLRKRSFGENGILDIIVEINLTLDIMEGASAPAEALVSESVRRRAATGICLTPLRKRSFGENGILDIIVEINLTLDIMEGASAPAEALVSESVRRRAATGICLTP15433..17108

/gene="C47G2.2"

join(15433..15770,16205..16531,16772..17108)

/gene="C47G2.2"

/note="Similar to fork head domain protein: cDNA EST EMBL:T01618 comes from this gene; cDNA EST EMBL:D70725 comes from this gene; cDNA EST EMBL:D66805 comes from this gene; cDNA EST yk268f8.3 comes from this gene; cDNA EST yk268f8.5 comes from this gene; cDNA EST yk474a11.5 comes from this gene; cDNA EST yk637f8.3 comes from this gene; cDNA EST yk670a12.3 comes from this gene"

/codon_start=1

/protein_id="CA88935.1"

/db_xref="GI:3875031"

/db_xref="SPRTEMBL:Q18694"

/translation="MPEMSSTLSSTPKLEPPPKLEPPETINQVYDRLSRTRLEPSTISAVLEHDLKFGESKRKRSIGDPTEDDGVPRKANKRNHSTSSADSSDAKDQDDDDSTSRKSGSRKRSNAKRPYSIALIAMSILNSPEKRLISECEPIIKFEYKKEFPAMONSIRHNLSDCEVAVARPGPCGNGWALDPDNEHMFNSFLRRKRYKKNSPDTHSMHPPRPDPGPPRRMHPRANPLMGCHMPPRAPVNMFAFIPOINTDOKLLSMASRIIMPAPVSSGGRKSSSSSPRENNSSAVSDKLSA"complement(24417..26914)

/gene="C47G2.4"

complement(join(24417..24975,25043..25176,25237..25522,25668..25907,25967..26054,26106..26227,26277..26608,26681..26785,26846..26914))

/gene="C47G2.4"

/note="Similarity to Trichostrongylus colubriformis 11 kd secretory protein (Swiss Prot accession number P21937); cDNA EST EMBL:D33349 comes from this gene; cDNA EST EMBL:D37644 comes from this gene; cDNA EST EMBL:D6149 comes from this gene; cDNA EST EMBL:C11456 comes from this gene; cDNA EST EMBL:C13631 comes from this gene; cDNA EST yk358b10.3 comes from this gene; cDNA EST yk358b10.5 comes from this gene; cDNA EST yk301f4.3 comes from this gene; cDNA EST yk301f4.5 comes from this gene; cDNA EST yk295f2.3 comes from this gene; cDNA EST yk295f2.5 comes from this gene; cDNA EST yk550d1.3 comes from this gene"

/codon_start=1

/protein_id="CA88936.1"

/db_xref="GI:3875032"

/db_xref="SPTEMBL:Q18695"

/translation="MGTISLAVOLFIEFLTSYLKNTIRKONPIVTISTFGYFSLIVFLPDLVATTPFKKENDRORILNNTTSAPRAVVECELPGGVYDDVFLNMRVYVNSQLTWLPILOSQVYTAGNTILGKITAAVYNNALYATISLCLFALITIAMKGVSTINLEKLVIVSASVTWGLFVNLVGLGDELPRSLHNGNRIRLAKTFFDIERLASEKSEAEENVADIKYKRVALLNSKNGORRVRTILSFSDVDIDNLEP SRQYIDNAHLDESQSEAKLISLHKRTIYAVQILNNAQMKVLDRAFLFLNLAIESNGYNLELSDPVGVRRENTYRQTEPCRLGIVTEVMEFVLSCEFEVYVYTLSPAFTVYASPRHYKYTQPAFGIIVLYITSAYETIIFRLOIYVLLDNGHDENSLIESAILLCRTPIPCNLNGLHMHDSHISNAKSEFGLEOTFKMGHLYPIPLAKGNIYLPICILICATIHRYVGAVYLNHIGFDQFVADETNDMINSGRSLVOIERNSIKRNSRDSORTOMTNSFGSSNAGSTTSKRKRNKNDDEERLEDDDEVEESSTRISMSPTHEPSSSGFPDDM"

30076..34049

/gene="C47G2.5"

join(30076..30134,30184..30492,31050..31274,31321..31480,31534..31653,31701..31835,31889..32104,32160..32570,32639..32935,32998..33215,33263..33651,33802..34049)

/gene="C47G2.5"

/note="cDNA EST CEMSB75F comes from this gene; cDNA EST EMBL:C13621 comes from this gene; cDNA EST yk227d8.5 comes from this gene; cDNA EST yk257e7.5 comes from this gene; cDNA EST yk319c4.5 comes from this gene; cDNA EST yk319c4.5 comes from this gene; cDNA EST yk335d12.5 comes from this gene; cDNA EST yk486d7.5 comes from this gene"

/codon_start=1

/protein_id="CA88937.1"

/db_xref="GI:3875033"

/db_xref="SPTEMBL:Q18696"

/translation="MAAEGKRNNRKRKSRDRREGTSASDAYSSDSADDEDITQDDDDPMLNLDDSDHNVVAPRQEPDPLIVISPNMEMPDNEDLEHEMKPEEGSGSR IADSEASDVTGDSRSDSDSHTRASLCYHRKSNMSPAKSLMPTISIPMFAKDEE ENSLMRLKTDNFTLIDVLLNLFVYQSRGKAELVQYITSRNMKALLETSLNPKTN TDLPMKQOURLSTIASETLITRTDVFQKQIVTTEBTRKLVPLNDKPTLNLVAGF FAFIMCELLSRHFDLTFTSLQETKRFDKCLRNINLGAIECLLENLIVRIPITSSEGRIV

				KEMHISENLEPKYIDRRARESTDDKEKELAVYEIILRELDKLYIMESKYDELJHAKS MDEFLIAKIADNLIVBEGCPAEEELVKRSALISASAKTILEATIKTNFVSNAPOULEET ERNLIEREHSYSGLMRPMCDMPYEHSDPPEIRVDSILANRLPNLIQTVLRIDEAN GSWQPLRIILIELCNTNCMSHTERKIVAEPSELPFINLIKAKMLPRASYHLCLLVK VILLIHSEFPCEDELSPAAYEYLITGEGGLONTIYDPTSPNPSSVASCISRFNONG AINRAKRAGIENOKLUALISADNTWTTEEDIILHYLNKHRPOMOHDFENDSVSIRS DSHFNDSEEWTDASTRKPAEMDATNSKAOCSGSPREPMMORSHOPECOFDPPDE DEFKLCSERANSSSCAGISIFETPIPIKWGAERTSEKASPVPVASTYFOQTNGCN AFILOEBSGDGEWMPVTPPLETELVTYOTGRHPNNMWDETAPFSLHDMAPEEDDW WADPSSEPTTISPRTAAANSASSSSSDAMGSDIHLQGEASDWPPLNNSHSKASDEVMOG LAASISHPGDSSEA"
BASE COUNT	10589 a	6786 c	6432 g	11024 t
ORIGIN				
Query Match	18.2%	Score 323.8;	DB 34;	Length 34831;
Best Local Similarity	74.2%;	Pred. No. 2.1e-46;		
Matches 409; Conservative	0;	Mismatches 142;	Indels	0;
Gaps	0;			
Oy	496	taccagaattgacgtcgatatacgcgatcaacaagcttccatggaaagcgtcataaaaagt	555	
Dd	3883	TTCCTAAAATTTCACAGCGCCTACCAGTATCCAAATGCTTTCAATAGAGTCCGACAAGACTGT	3942	
Oy	556	tagtcacagaatttgaggatctcgaatatcaacaactgctttccaactcaaatctgcocat	615	
Dd	3943	GMCCACAGATCGAAGTGCTCGAGTCTCACACCTGCCTCCAGACCACGGTGTAACCAT	4002	
Oy	616	gcccatagccgcttatgtaattttgtagtygbyggaccaacccggtcaaccaagtccaattgc	675	
Dd	4003	GCCAGTCTGCAAATACGAGATCTCTTGACGGAGGACCATTCGGACAAACCATCAATGTCG	4062	
Oy	676	tatcatgtgcacccagattatcatataaatygaacaycgatcttcgaaacctgttacatt	735	
Dd	4063	CACCATCGGACAAAGAAGTCTATCACAAATGACCTTGCGATTGTGAGACCATGACACCTT	4122	
Oy	736	ctcgcgagttgtccaatctcctgttgtcgatgayaigtlaaagtgatctatcytggaaattct	795	
Dd	4123	CTGCCCGCTCGTTCACCTCTTGACACGTGTGATGATGTATAAGCGAACCCGTTCAATCTT	4182	
Oy	796	aaatctatgatatgtcgtcttgataaataatttgtcaataataattggaaatcacacaaga	855	
Dd	4183	TACGACAGAGAGATGTGCTCTTGACAAAGTCTTGTGCTCAATTAACCTTGGAGTACCCAACTGA	4242	
Oy	856	tttaa'tgctgtgcacaagaagctacgylabacaataatgcgaatcatcaacagctttcta	915	
Dd	4243	CTTGATGCTGCGCCCAAGAAGCTCAGGTCTACAAATATGCCGATCGCTCCCACTCTTCTA	4302	
Oy	916	tcaatgccaagatcaattatcatataaagaaccaaatgagaaattglttgcagcaaatg	975	
Dd	4303	TCAATGCCAAATCTCCATCATCCATCAAGAACCCAGAGAACGATGGCCGCTCCAATTGG	4362	
Oy	976	ttcaagaaccacaagaattctcgagcgtttaaacaagtgtygtcgcgaagcaaaacctgtgc	1035	
Dd	4363	CTCAGAGCCACAAAGATTGCGAGCGGTCAACAAAGGTGTTGCCGGAGGAGCTCATAGCCGC	4422	
Oy	1036	agctgcgcgaac	1046	
Dd	4423	CGCTGCTCCAC	4433	
RESULT 3				
CELCTU1				
LOCUS	CELCTU1	2310 bp	DNA	INV
DEFINITION	C.elegans cuticlin 1 (cut-1) gene, 3' end.			
ACCESSION	M55997			26-APR-1993
VERSION	M55997.1 GI:156271			
KEYWORDS	cuticle protein; cuticlin 1.			
SOURCE	C.elegans DNA.			
ORGANISM	Caenorhabditis elegans			
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderiinae; Caenorhabditis (bases 1 to 2310)				
Sebastano,M., Lassandro,F. and Bazicalupo,P.				
AUTHORS				

TITLE	CUT-1 a Caenorhabditis elegans gene coding for a dauer-specific noncollagenous component of the cuticle
JOURNAL	Dev. Biol. 146, 519-530 (1991)
MEDLINE	91323673
FEATURES	
SOURCE	location/Qualifiers 1..2310
	/organism="Caenorhabditis elegans"
	/strain="N2"
	/db_xref="taxon:6239"
mrna	join(1186..1920,1971..2310) /partial
	/gene="cut-1"
	/product="cuticlin 1"
	1186..1920
exon	/gene="cut-1" /number=2
	/product="cuticlin 1"
gene	join(1186..1920,1971..2310) /gene="cut-1"
	join(1186..1920,1971..2162) /partial
CDS	/gene="cut-1" /codon_start=1 /product="cuticlin 1" /protein_id="AA27995.1" /db_xref="gi:156272" /translation="RIPPCPYMSDEKTSVNIQVSDLTFAFOTYVPMPVKYEILDGSPSGPIQFATIGQVYHKKTCDSERTDFCAAVHSCTVDGNGDIYVLNEGALLDKPLINDLEYPTDMDGOEAHVYKVAORSOLFQCQISTIKDGSSCARTCSEPGFGAARQAAGAGAHAANAPOAGVEVOAPAAGAAPVAAAAPVAPRTLAQLLRRKRSGENGGIILDVREIVEINTIDIMGASPSAPEAALVSESVRRRASTGTISRPGLSPFLGMRTIVATLSATIFYVARPTSKHK" 1971..2310
exon	/number=3 /gene="cut-1" /product="cuticlin 1"
BASE COUNT	642 a 567 c 433 g 668 t
ORIGIN	
Query Match	17.7% Score 315; DB 34; Length 2310;
Best Local Similarity	75.1%; Pred. No. 9.1e-45;
Matches 393; Conservative	0; Mismatches 130; Indels 0; Gaps 0;
OY	524 caatgctttaaatgaagcgtatataaaacagttaatgcacagattgaggatatcgaatc 583
Db	1198 CCATCTTCTACATTGAGAGTCGACAAGACTGTGCACACAGATCGAGTCCGATCTC 1257
OY	584 acaactgctttcaactcaaatgtcccgatgcagtatgcgctlatrgaaatttggat 643
Db	1258 ACCACTTGCTTCCAACCACCCAGTCTCTACCAAGCCAGCTGCAAAATACAGATCCTTAC 1317
OY	644 ggtagcaccaaccgycgacaaccagttcaatttgtcatcttgatcgacgagttatcataa 703
Db	1318 GGAGGACCATTCGGGACCAACCAATTCAGTTGCCACCATTCGGACAACAAGTCTATACAAA 1377
OY	704 tggacatgcgatctctgaaacgcttgatacttcctgcgcygtgtlccatctcgtcttg 763
Db	1378 TGGACTTGGAGATTCTGAGACCACTGACACACTTCTGCGCGGTTCACCTTGCACTGTC 1437
OY	764 gatgtagtlaaogtgcatactgtggaattctcaaattgctatgtagtgcgtcttgataa 823
Db	1438 GATGATGTAAGTGAAGGACCCGTTCAAGATTCTTAAGAAAGAGATGTGCTTGACAAAG 1497
OY	824 taattgctaataaatttggaatalatccaacagaattaaatgctgcgcgaagaagctcacgta 883
Db	1498 TTCTTGCTCAATAAATCTTGAGAGTACCAACTGACTGATGAGGCTGGCCAAGAGCTACGTC 1557
OY	884 taacaatatgcgatacgatacacagctttctatacgaatgcccagatcagatlaaccaataa 943
Db	1558 TACAATATGCGCGATCGCTCCCAACTCTTCAATCAATCCCAATCTCCATCACCATTCAC 1617
OY	944 gaaccaataaacgaatgcatactgacctacacatttbaagaacacacgaagaattgaacatft 1003

```

Db      1618 GACCAGGAGGAAGCAATGTGCCCCCTCCAACTTGCTCAGAGCCACAAAGATTGCGAGCCGTC 1677
Oy      1004 aaacagctgtgtcgcagcaaaaacctgtcagctgcgcgcac 1046
Db      1678 AAACAAGCTGTGTCGGAGAGAGCTCATGCGCGCTGCTCCAC 1720

RESULT  4
AF125580 724 bp mRNA INV 05-MAY-1999
LOCUS    Wuchereria bancrofti cuticulin-1 mRNA, partial cds.
DEFINITION
ACCESSION AF125580
VERSION    AF125580.1 GI:4741874
KEYWORDS
SOURCE    Wuchereria bancrofti.
ORGANISM  Wuchereria bancrofti.
Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;
Filarioidea; Onchocercidae; Wuchereria.
REFERENCE 1 (bases 1 to 724)
AUTHORS   Ramzy,R., Helmy,H., Adely,M., Curtis,K. and Weil,G.
TITLE     Wuchereria bancrofti L3 cuticulin-1 cDNA partial sequence
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 724)
AUTHORS   Ramzy,R., Helmy,H., Adely,M., Curtis,K. and Weil,G.
TITLE     Direct Submission
JOURNAL   Submitted (03-FEB-1999) Research & Training Center on Vectors of
Diseases, Ain Shams University, Abassia Square, Cairo, Egypt

FEATURES
Source    location/Qualifiers
1..724
/organism="Wuchereria bancrofti"
/db_xref="taxon:6293"
/dev_stage="L3 larvae"
/country="Egypt; Nile Delta"
<1..623
/note="similar to Ascaris lumbricoides cuticulin-1"
/codon_start=3
/product="cuticulin-1"
/protein_id="AAD28743.1"
/db_xref="GI:4741875"
/translation="TVDFECAVHSCFVDDGDKVELLNADGALDKFLINNEYPT
DLMGQEAHVYKADRSOLFQOCOSTIKRPNKQCPQCTEPOGFAITRNQVAA
ASRQAAFRVLKRDYRENTVDVTDLNADINEEFLSPALHRSLSLAHENGHP
VIVATGTGICMSTGTFTLAKLILFVIVSVATVATVATILLRSHSTKV"

BASE COUNT 248 a 112 c 128 g 236 t

Query Match 14.8%; Score 263.4; DB 35; Length 724;
Best Local Similarity 62.4%; Pred. No. 6,4e-36;
Matches 449; Conservative 0; Mismatches 261; Indels 9; Gaps 2;

Oy      720 aaacgttaactcttcgcgcgtgtcattccctgcttgatgtagtgaacgtgtg 779
Db      1 AAACAGTTATTCATTCTGTGCGGTGATTCATTGCTTTGTGATGATGGTAATGGTG 60

Oy      780 atactgtggaatatcctaattctgtagtgcctcttgataataatatctgctaataatc 839
Db      61 ATAAAGTAGACGTATTAATAATGCTGATGTTGTGCTTTGGACAAATTTCTACTTAATAAT 120

Oy      840 tggataatcacaagaatcagctgcgcgaagaagctcagctatacaataatgcgagatc 899
Db      121 TAGAATATTCACAACAGATTAAATGGCTGGACAAGACACGTAATACAAATCGCTGAAC 180

Oy      900 gatacagcttctatacatgcagatcagatattaccattaaagaaccaaatagcgaat 959
Db      181 GATTCACACTTTTATTCATATGTCATAATTAGCAATTAAGGACCAAAATAGTAAT 240

Oy      960 gttcttcagacacaatgttctagaacaagaatctcgaagctttaaacagc---tgctg 1016
Db      241 GTCAAGGACCAATGACTGAACGCAAGGTTTGGCGCAATTAACGCGCATGTGG 300

Oy      1017 ccgacgaaaaaacctcgtcagctgcgcaacttcgttactcaagaagaatcgtcagaac 1076

```

```

Db      301 TTGCAGCTGCATCAGACACAAGCTGCTGCAATTAGAGTATTGAAAAAGAGATGTAAAG 360
Oy      1077 cggagaatcattatgatragsaacctgatatacaaacctttaaattgcgtgtaac 1136
Db      361 ATGAATAATATGTGTGTGACGTACCGACTTAAATGATGATTAATGAAGAGTTTA 420

Oy      1137 aagcttgcagctgattacgtcacgcgtgcactctg-----caacataatgagcaac 1190
Db      421 CCTCATTTACCAATAGCGCTACGTCATCGATCATCTTTGTTACACATGAGATGGCTCATC 480

Oy      1191 ctgtaatactgtcgcagtagacaanaatggaatcgtcagctaccacttgcgttcaatgt 1250
Db      481 CGGTAAATTTGTCAACAATGACACAAAGGATATGTGATGTCATTAACCGGTTTACACTTG 540

Oy      1251 ttatggtttaagcaatgatgatgtgcgcgcgtatcatatcatccttgcgttaattc 1310
Db      541 CCGGTAATGCTAATATTTATTTGTTCTGTGCTACATATGTTGCTATTTACATTGTTAC 600

Oy      1311 gccaatcagaagcataaaataatgltgaatcatcgaagcaataaataactgcga 1370
Db      601 GATCACATTTCTACCAAGTTTAAATGTTATTAATGATATTACACATCTATGTTATTC 660

Oy      1371 tatatactgcttctcttatacatcctcctaataatattgactaacaatatata 1429
Db      661 ACTATTGCTAGCTCTTTTCTTCTTATTTGAACCTTAATTAATTAATTAATTAATTAATTA 719

RESULT  5
ALU073005 5792 bp DNA INV 14-AUG-1997
LOCUS    Ascaris lumbricoides CUT-1-like cuticlin protein precursor
DEFINITION (ascut-1) gene, complete cds.
ACCESSION U73005
VERSION    U73005.1 GI:1657624
KEYWORDS
SOURCE    common roundworm.
ORGANISM  Ascaris lumbricoides
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Ascarididae;
Ascaridoidea; Ascarididae; Ascaris.
REFERENCE 1 (bases 1 to 5792)
AUTHORS   Timinouni,M. and Bazzilalupo,P.
TITLE     cut-1-like genes of Ascaris lumbricoides
JOURNAL   Gene 193 (1), 81-87 (1997)
MEDLINE   97390131
REFERENCE 2 (bases 1 to 5792)
AUTHORS   Timinouni,M. and Bazzilalupo,P.
TITLE     Direct Submission
JOURNAL   Submitted (01-OCT-1996) GA3, I.I.G.B., via G. Marconi, 10, Napoli,
NA 80125, Italy

FEATURES
Source    location/Qualifiers
1..5792
/organism="Ascaris lumbricoides"
/variety="sum"
/db_xref="taxon:6252"
766..881
/gene="ascut-1"
/join(766..954,2851..3239,3583..4083,4499..5081)
/gene="ascut-1"
/product="CUT-1-like cuticlin protein ASCUT-1"
766..5081
/gene="ascut-1"
766
/gene="ascut-1"
/note="site of transplicing by the SL1"
882..929
/gene="ascut-1"
/join(882..954,2851..3239,3583..4083,4499..4693)
/gene="ascut-1"
/note="cuticlin gene: ASCUT-1; homologous to cuticlin
proteins of other nematodes"
/codon_start=1
/product="CUT-1-like cuticlin protein precursor"

```


Oy	397	tcatttgatcatcgcaatgctgttgagctaaacgatctcgaaatccacgysgfatllttgt	456
Oy	397		456
Oy	397	tcatttgatcatcgcaatgctgttgagctaaacgatctcgaaatccacgysgfatllttgt	456
Oy	397		456
Oy	709	tccggttgacactcttgcatatgttgccccacatgatgct-----gaaccacaagagatgttctg	765
Oy	457	aacaacaactgltgcatlcttcgtlccatccatatttgttaccaaaagtgtgacgtgcata	516
Oy	457		516
Oy	457	aacaacaactgltgcatlcttcgtlccatccatatttgttaccaaaagtgtgacgtgcata	516
Oy	457		516
Oy	766	ctccacacacggctgctatcttccatctccatccacagttttgtaccacaagtgtgaccgggacct	825
Oy	517	tcagatacaatgcttcttaccatggaagctgtgataaacaagttgaatgacaaatggaatgc	576
Oy	517		576
Oy	517	tcagatacaatgcttcttaccatggaagctgtgataaacaagttgaatgacaaatggaatgc	576
Oy	517		576
Oy	826	ccggttcacagctcttcatatgaggaagacaaagaccgtctcgccacacactcgacagtgctc	885
Oy	577	tgaataccaactctgtcttcaaacatcaaatgtcccgatgacagtaatgacgtgacgtgac	633
Oy	577		633
Oy	577	tgaataccaactctgtcttcaaacatcaaatgtcccgatgacagtaatgacgtgacgtgac	633
Oy	577		633
Oy	886	cgagatctcacacacagattccacagaccagtggtgctctatgctgtotgcacaaatgatgag	945
Oy	633	-----	633
Oy	633	-----	633
Oy	946	aagtagaagaagaagac	1005
Oy	633	-----	633
Oy	633	-----	633
Oy	1006	aaatcgatgctgatttaatttcacattttgcttccattaccacacgtttaagatgctcaaa	1065
Oy	633	-----	633
Oy	1066	aatggtttgtttttaacgttttatgtataaaaaaacaaacaaatattttttaatgaatt	1125
Oy	633	-----	633
Oy	633	-----	633
Oy	1126	tatttttcacaaagatccttcgaagccgcccacacttgggcacaaccacatttcacattgc	1185
Oy	682	tgttcagccagttatcatataatggaatgcagatctctgaaacagttgacttctgcgc	741
Oy	1186	tgcgcacacaaagtctacacaaagtggagactgtgactctgacacgattgacactttttgcgc	1245
Oy	742	ggtgttcacatcctgcttgcgtgcatgatagtgataacggtgatactgtgaaatcttaatgc	801
Oy	1246	ggtgttcacatcctgcttgcgtgcatgatagtgataacggtgatactgtgaaatcttaatgc	801
Oy	802	tgatgagatgagctcttcatataatattgtcacaataatttggaatatccaaagatttat	861
Oy	1306	aggaagctgtgctccttgacaaagttcttgctcaacaacttggaattaccacacgatttgac	1365
Oy	862	ggtctgcacaaagaactcaacgtatatacaaatatgagatcgatcaacagcttcttatacaatg	921
Oy	1366	tgcgcggccaaagagctcagctctacaaatattgcggagattgcctctcaaacctttttaccatg	1425
Oy	922	ccagatcagttatccatataagaaccaaataagcgaaatgltgttgacacaaatgltcaga	981
Oy	982	acacaaagatctggaagctgtttaaacaagtgatgctgcagacaaacacgtgcagctgc	1041
Oy	1486	gccaaagccgcttcaacccctcaaaagtgggcgcggctgagctggtgcccgggtccacac	1545
Oy	1042	gc 1043	
Oy	1546	gc 1547	
RESULT	7		
LOCUS	BMA012617	637 bp DNA	INV
ACCESSION	AF012617.1	GI:3858954	29-JUN-1999
VERSION	AF012617.1	GI:3858954	
KEYWORDS	cut-1 gene; cuticlin.		
SOURCE	Brugia malayi.		
ORGANISM	Brugia malayi		
	Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spiruridae;		
	Filarioidea; Onchocercidae; Brugia.		

REFERENCE		1 (bases 1 to 637)			
AUTHORS	Lewis,E., Hunter,S.J., Tetley,L., Nunes,C.P., Bazzicalupo,P. and Devaney,E.				
TITLE	cut-1-like genes are present in the filarial nematodes, Brugia pahangi and Brugia malayi, and, as in other nematodes, code for components of the cuticle				
JOURNAL	Mol. Biochem. Parasitol. 101 (1-2), 173-183 (1999)				
MEDLINE	99339397				
REFERENCE	2 (bases 1 to 637)				
AUTHORS	Devaney,E.				
TITLE	Direct Submission				
JOURNAL	Submitted (06-NOV-1998) Devaney E., Veterinary Parasitology, University of Glasgow, Bearsden Road, Glasgow, G61 1QH, UK				
FEATURES	Location/Qualifiers				
SOURCE	1. 637				
	/organism="Brugia malayi"				
	/db_xref="taxon:6279"				
gene	/clone="cut-1"				
	1..484				
exon	/gene="cut-1"				
	1..164				
	/gene="cut-1"				
	/number=1				
mRNA	join(<2..164,266..>484)				
	/gene="cut-1"				
CDS	join(<2..164,266..>484)				
	/gene="cut-1"				
	/codon_start=1				
	/product="Cuticlin"				
	/protein_id="CAAI0074.1"				
	/db_xref="gi:3858955"				
	/db_xref="sp trEMBL:O96775"				
	/translation="VISPHPLSVTKVDKAVYVOCFYMEADKTYSNIEVSITTAFQI GIWPCRYVELLNGSGPGIOPIATISGPVYHHMTCDSIEIVDFCAVHSCFVDDG GVPEILISADGCALDKYLINLEP"				
	163..265				
intron	/gene="cut-1"				
	/number=1				
exon	266..484				
	/gene="cut-1"				
	/number=2				
BASE COUNT	174 a 136 c 134 g 193 t				
ORIGIN					
Query Match	10.7%; Score 190.6; DB 34; Length 637;				
Best Local Similarity	68.3%; Pred. NO. 1.6e-23;				
Matches 334; Conservative 0; Mismatches 54; Indels 101; Gaps 1;					
QY	469 tgcatttcgcttcatcattatcttgtttaccaaagtgatcgtagatacgaagtacaag	528			
Dd	1 tgtaatttccctccawccacgtccgtccgcgcacaaaaggttgatcgacgactccgagtcag	60			
QY	529 ctcttacatggagaagctgataaaacagctgaagtcagatgcaagatgatatgaaatacaaac	588			
Dd	61 cttctaattatggagaagctgacaaaacggtagagccacagattgagtgctgaanaatcacAAC	120			
QY	589 tcgctttaaacctcaaatctgtcccgatgccagatcgcgcgttatgaa-----	635			
Dd	121 tgcctttccaacctcaaatgttcccattgccctgtttgtgatgacgagcatgcttttttga	180			
QY	635 -----	635			
Dd	181 atttttgcatttgcattttattatACAATTCAATAACATGAANAATACTATTATCA	240			
QY	635 -----atttgatggttggaaccaacggtcaaccagtt 667				
Dd	241 CATAGTAGATGGTGACAATAATTTTGTAGATTTTGGATGGTGACCACAACCGGAGACGCCAATT	300			
QY	668 caatttgatcatcatgtgtcagccagcttatcatctaaatgacatgagatctcgtgaacagctt	727			
Dd	301 CAATTTCCTACCATTTGGCCAAACCAATTATATCAACAATGAGACTGTGATTCGGAACCGGTT	360			

```

Oy 728 gatacttcctgcgcgggtgtccatctcgtcttgatgatagtgaacggtgatactctg 787
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 GATACCTTCTGCGAGTGTGCACCTGCTTGTGATGATGACGAGTGATACGGTGTG 420
Oy 788 gaattcctaattcgtcgtgatagtgctcttgataataatttgctaaataattgaaatat 847
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 GAAATTCGAGTGCAGAGTGTGTCTCTCGATTAATACTTGCAATTAATTTGGAATAT 480
Oy 848 caacacgat 856
      ||||| |||||
Db 481 CCGACAGGT 489

RESULT 8
BPA012618 358 bp DNA INV 29-JUN-1999
LOCUS DEFINITION Brugia pahangi cut-1 gene, partial.
ACCESSION AJ012618
VERSION AJ012618.1 GI:3858956
KEYWORDS cut-1 gene; cuticle.
SOURCE Brugia pahangi.
ORGANISM Brugia pahangi
Eukaryote; Metazoa; Nematoda; Secernentea; Spirurida;
Filarioidea; Onchocercidae; Brugia.
REFERENCE 1 (bases 1 to 358)
AUTHORS Lewis, E., Hunter, S.J., Tetley, L., Nunes, C.P., Bazzicalupo, P. and
Devaney, E.
TITLE cut-1-like genes are present in the filarial nematodes, Brugia
pahangi and Brugia malayi, and, as in other nematodes, code for
components of the cuticle
MOL. Biochem. Parasitol. 101 (1-2), 173-183 (1999)
JOURNAL 993939397
MEDLINE 2 (bases 1 to 358)
REFERENCE Devaney, E.
AUTHORS Direct Submission
TITLE Submitted (06-NOV-1998) Devaney E., Veterinary Parasitology,
University of Glasgow, Bearsden Road, Glasgow, G61 1QH, UK
JOURNAL Location/Qualifiers
FEATURES
source 1..358
/organism="Brugia pahangi"
/db_xref="taxon:6280"
join(<1..191,350..>356)
gene mRNA /gene="cut-1"
exon 1..191
/number=1
gene /gene="cut-1"
1..356
/number=1
CDS /gene="cut-1"
join(<1..191,350..>356)
/note="unnamed protein product"
/codon_start=1
/protein_id="CAA10075.1"
/db_xref="GI:3970656"
/cd_xref="SPTRMBL:096776"
/translation="PIGFATIGPQVYHKMTCDSEVTDFCAVHSCFVDGNGDTVEI
LNAGCALDKLYLNNLEPYTDL"
intron 192..349
/gene="cut-1"
/number=1
exon 350..356
/gene="cut-1"
/number=2
BASE COUNT 110 a 64 c 64 g 120 t
ORIGIN
Query Match 8.8%; Score 156.6; DB 34; Length 358;
Best Local Similarity 87.7%; Pred. No. 1..1e-17;
Matches 171; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
662 ccagttcaattgtctatcattgtgtcagccagttatcataaatgacatgcgattcttga 721
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 1 CCAATTCATTTGCTACCATTTGGCCACACAGTTTATACAAATGACCTGTGATCCGAA 60
Oy 722 accgttgatactcttcgcgcgggtgtccatctcgtcttgatgatagtgaacggtgat 781
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 ACCGTTGATACCTTTTGGCGCACTTGTCCACTCTCTTGTGATGATGATGCGAACGGTAT 120
Oy 782 actgtggaattcctaattcgtcgtgatagtgctcttgataataatttgctaaataattg 841
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 ACGGTGAAATTCGATGACAGATGTTGTGCTCTCGCAAAATFACTTGCTGAAACAATTTG 180
Oy 842 gataatccaacgat 856
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 GAATATCCACACAGGT 195

RESULT 9
CEF53F1 39478 bp DNA INV 02-SEP-1999
LOCUS DEFINITION Caenorhabditis elegans cosmid F53F1, complete sequence.
ACCESSION Z81088
VERSION Z81088.1 GI:1627965
KEYWORDS HMG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilitia; Rhabdilitia;
Rhabdilitia; Rhabdilitia; Rhabdilitia; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 39478)
AUTHORS Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berts, M.,
Bonfield, J., Burton, J., Connell, M., Copey, F., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kersey, J., Kirsten, J., Laister, N.,
Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, A.,
Saunders, D., Showkneen, R., Smalton, N., Smith, A., Sonhammer, E.,
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Sproat, J. and Woldman, P.
TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2 (bases 1 to 39478)
AUTHORS Burton, J.
TITLE Direct Submission
JOURNAL Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail:
jes@anger.ac.uk or rwenematode.wustl.edu
COMMENT Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
For a graphical representation of this sequence and its analysis
see: -
http://webace.sanger.ac.uk/cgi-
bin/display?db=wormacexclass=Sequence &object=F53F1
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
This sequence is the entire insert of clone F53F1. The true right
end of clone M04612 is at 21759 in this sequence. The start of this
sequence (1..101) overlaps with the end of sequence Z81103.
The end of this sequence (39379..39478) overlaps with the start of
sequence AL021448.
location/Qualifiers
1..39478
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="V"

```

```

gene
/clone="F53f1"
complement(14923..16486)
/gene="F53f1.1"
CDS
complement(join(14923..15347,15663..15774,15835..15984,
16033..16271,16318..16486))
/gene="F53f1.1"
/feature="predicted using GeneFinder; similar to cuticlin"
/codon_start=1
/protein_id="CAB03124.1"
/db_xref="GI:3877456"
/translation="MDNKEIRRTLEDYTPKGFLLSESSKYDCHLEDMLVQLLLATL
AVNPVIONSLGDVQVCDSTRTISVQIKTEPPEGVLFVDFASEEVCSTGRRLS
AFLEIEIGALRQVRLPKGLAVRTTITSEHPTTKVDRYTNLLCLYRESQVTS
ANNISVDEISRTISYVNLMPCTYQIILSGFPGPVEFGIGQGVYHQMKODNKGK
IILIKEDSCVAVHTGSCVDGKSTGLSDSGCSIDKFLSLNLEYPCNLLAGOPAHY
YKFAADRALFROCCISITVKEPDCVAPICEDVGGAPVAVPPPPYMIKSNLVNIN
OKNSFTPEKSPMYVVCITNHNKAIOKL"
17396..18524
/gene="F53f1.2"
join(17396..17599,17683..17926,17977..18146,18197..18286,
18339..18524)
/gene="F53f1.2"
/feature="similar to aldehyde reductase; cDNA EST yk473d8.3
comes from this gene; cDNA EST yk473d8.5 comes from this
gene"
/codon_start=1
/protein_id="CAB03127.1"
/db_xref="GI:3877459"
/db_xref="SPTREMBL:P91997"
/translation="MSDSFANVPGSSQOLNSGYNIPYVEFGTYKTKGNPPAIDAL
TAGYRMPDTAKYVLYNEKELEGALKLLPKHRSRDVLTSEKFPESNACGACGV
ESLQSLQTDIDMYLVHPRKNDNDVNAEYRKIAYLEBKAAGVYRSIGVS
NVEIYHELEKTYAKVPCANOLEYHPHARIPLOKYCKERNIFEQASSLARHEPKL
IEDPVVLELAKKHNTPVPLVLLMALRQNVGIVPKSVPSRIVENKVIDIALRPEDI
OSLTLDDGQHYIRCTGWLK"
18714..20251
/gene="F53f1.3"
join(18714..18887,18935..19056,19113..19302,19353..19448,
19519..20008,20072..20251)
/gene="F53f1.3"
/feature="similar to aldehyde reductase"
/codon_start=1
/protein_id="CAB03128.1"
/db_xref="GI:3877460"
/db_xref="SPTREMBL:P91998"
/translation="MSIKLNTGYDCPLGLCTYKIIQDQVLPVDAALTAGYRLDPA
KYVNEKEIGALETILPKHNLKREDITTKMHPNTEVNAKVLDESLSILKTSYID
MYLIHYKSPDYGDPMNKTLRIATNMDWECKNAGKIRSGVSSPEIRRLLEKQDL
GKNPPCCNOVEYHHTREELKNTCKSEGIFFOAFSSLAHNETILLESITTRIAEK
IYHPRKTVLLSMAISQKVGIIIPKSTNPERLQNLTVLLEBEVAKKICNLNDKHYYR
TTGWLVL"
21582..22009
/gene="F53f1.4"
join(21582..21670,21781..22009)
/gene="F53f1.4"
/feature="predicted using GeneFinder; similar to cuticlin;
cDNA EST EMBL:D72018 comes from this gene; cDNA EST
EMBL:D74799 comes from this gene; cDNA EST EMBL:C11739
comes from this gene; cDNA EST EMBL:C12451 comes from this
gene; cDNA EST EMBL:C10189 comes from this gene; cDNA EST
EMBL:C10682 comes from this gene; cDNA EST yk506b3.3 comes
from this gene; cDNA EST yk437e9.3 comes from this gene;
cDNA EST yk437e9.5 comes from this gene; cDNA EST
yk311h10.3 comes from this gene; cDNA EST yk311h10.5 comes
from this gene; cDNA EST yk341f6.3 comes from this gene;
cDNA EST yk341f6.5 comes from this gene; cDNA EST
yk300g2.3 comes from this gene; cDNA EST yk300g2.5 comes
from this gene; cDNA EST yk324b3.3 comes from this gene;
cDNA EST yk324b3.5 comes from this gene; cDNA EST
yk191d9.3 comes from this gene; cDNA EST yk191d9.5 comes
from this gene; cDNA EST yk677d4.3 comes from this gene"
/codon_start=1
/protein_id="CAB03125.1"

```

gene
CDS

```

/db_xref="GI:3877457"
/translation="MOKIVFPAIAVSOAFLPSGGCGGACAPPPEPPCCGCGPS
LPLQLPREFLPRLSLPSFGCGCGGAPGCGAPVAVAPPAGVTAAPPAGGYATA
GRK"
25572..26408
/gene="F53f1.5"
join(25572..25660,25850..26408)
/gene="F53f1.5"
/feature="predicted using GeneFinder; similar to cuticlin;
cDNA EST EMBL:T01107 comes from this gene; cDNA EST
CEESN60P comes from this gene; cDNA EST EMBL:D70966 comes
from this gene; cDNA EST EMBL:D71580 comes from this gene;
cDNA EST EMBL:D72120 comes from this gene; cDNA EST
EMBL:D72428 comes from this gene; cDNA EST EMBL:D73526
comes from this gene; cDNA EST EMBL:D74423 comes from this
gene; cDNA EST EMBL:D74402 comes from this gene; cDNA EST
EMBL:D74922 comes from this gene; cDNA EST EMBL:D75307
comes from this gene; cDNA EST EMBL:D76322 comes from this
gene; cDNA EST EMBL:C12136 comes from this gene; cDNA EST
EMBL:D69807 comes from this gene; cDNA EST yk505b11.3
comes from this gene; cDNA EST yk492b11.3 comes from this
gene; cDNA EST yk492b11.5 comes from this gene; cDNA EST
yk480h10.3 comes from this gene; cDNA EST yk480h10.5 comes
from this gene; cDNA EST yk445b7.3 comes from this gene;
cDNA EST yk445b7.5 comes from this gene; cDNA EST
yk440b8.3 comes from this gene; cDNA EST yk440b8.5 comes
from this gene; cDNA EST yk377b11.3 comes from this gene;
cDNA EST yk377b11.5 comes from this gene; cDNA EST
yk473g12.3 comes from this gene; cDNA EST yk473g12.5 comes
from this gene; cDNA EST yk459g4.3 comes from this gene;
cDNA EST yk459g4.5 comes from this gene; cDNA EST
yk482d10.3 comes from this gene; cDNA EST yk482d10.5 comes
from this gene; cDNA EST yk455c6.3 comes from this gene;
cDNA EST yk455c6.5 comes from this gene; cDNA EST
yk417a12.3 comes from this gene; cDNA EST yk417a12.5 comes
from this gene; cDNA EST yk414h1.3 comes from this gene;
cDNA EST yk401f4.3 comes from this gene; cDNA EST
yk401f4.5 comes from this gene; cDNA EST yk401e5.3 comes
from this gene; cDNA EST yk401e5.5 comes from this gene;
cDNA EST yk399e6.3 comes from this gene; cDNA EST
yk399e6.5 comes from this gene; cDNA EST yk392h1.3 comes
from this gene; cDNA EST yk392h1.5 comes from this gene;
cDNA EST yk391b10.3 comes from this gene; cDNA EST
yk391b10.5 comes from this gene; cDNA EST yk342c6.3 comes
from this gene; cDNA EST yk342c6.5 comes from this gene;
cDNA EST yk345c4.3 comes from this gene; cDNA EST
yk345c4.5 comes from this gene; cDNA EST yk307f5.3 comes
from this gene; cDNA EST yk307f5.5 comes from this gene;
cDNA EST yk258a5.3 comes from this gene; cDNA EST
EMBL:CO9046 comes from this gene; cDNA EST yk306e6.5 comes
from this gene; cDNA EST yk314h1.5 comes from this gene;
cDNA EST yk314h1.3 comes from this gene; cDNA EST
yk343g4.5 comes from this gene; cDNA EST yk375f1.5 comes
from this gene; cDNA EST yk404c1.5 comes from this gene;
cDNA EST yk447c9.5 comes from this gene; cDNA EST
yk447e5.5 comes from this gene; cDNA EST yk488e6.5 comes
from this gene; cDNA EST yk382a5.3 comes from this gene;
cDNA EST yk506h10.3 comes from this gene; cDNA EST
yk507f12.3 comes from this gene; cDNA EST yk518b6.3 comes
from this gene; cDNA EST yk530f12.3 comes from this gene;
cDNA EST yk536e1.3 comes from this gene; cDNA EST
yk537c8.3 comes from this gene; cDNA EST yk540g5.3 comes
from this gene; cDNA EST yk545b10.3 comes from this gene;
cDNA EST yk557b11.3 comes from this gene; cDNA EST
yk562h11.3 comes from this gene; cDNA EST yk564d11.3 comes
from this gene; cDNA EST yk566f10.3 comes from this gene;
cDNA EST yk596a10.3 comes from this gene; cDNA EST
yk596e11.3 comes from this gene; cDNA EST yk614f1.3 comes
from this gene; cDNA EST yk618d8.3 comes from this gene;
cDNA EST yk620c8.3 comes from this gene; cDNA EST
yk620d5.3 comes from this gene; cDNA EST yk622e4.3 comes

```

Query Match

6.0%; Score 107.4; DB 34; Length 39478;

Best Local Similarity	61.7%	Pred. NO. 1.5e-09	Mismatches	106	Indels	Gaps
Matches	171	Conservative	0			
Oy	723	ccgttgatacttctgtgcgcggtlgtccatctcctgtctgttcgagatgtaacggtgata	782			
Db	15350	CAGAAGACTCATTTCTGCATGTGATGTTTCAACACCTGTTCCTTTCACGATGACGACGAGGANA	15291			
Oy	783	cggtgaaattcctaatactgtatgatgtgctctgtgataaatttgcgaataatttgg	842			
Db	15290	CCTCAATTTCTCATTAACATCTGGAATGGGTCTCAATTTGATTAAGTTTCTATTGTGGAACCTTG	15231			
Oy	843	aataccacaagatttaattgtctgcccagaagatctcaagctacatatacagatgcgcatg	902			
Db	15230	AGTATCCGGGANAACCTTACCTCGCGGCAAGAAGCTCATGTATGATTAAGTTTGGCTGATCGAG	15171			
Oy	903	caacagctttctatacaatgcccagatcagatattaccattaaagaaccaatagcgaatgtg	962			
Db	15170	ATGCTCTTTTCTTCAGATGTCGATATATCAATTAACCGTCAAAAGACGACACGAGAAATGCG	15111			
Oy	963	ttcgaccacaatgtccagaacacacagaagatcggaagc	999			
Db	15110	TGAGACCAATATGTGAAGATGTGAAGAGCGGTGAGAC	15074			
RESULT	10					
CEEO4D5						
LOCUS	CEEO4D5	31536 bp	DNA	INV	02-SEP-1999	
DEFINITION	Caenorhabditis elegans cosmid E04D5, complete sequence.					
ACCESSION	Z66496					
VERSION	Z66496.1	GI:1041307				
KEYWORDS	HTG; Cuticle-In-Like protein.					
SOURCE	Caenorhabditis elegans.					
ORGANISM	Caenorhabditis elegans					
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilitida; Rhabdilitae; Rhabdilitidae; Rhabdilitae; Peloderinae; Caenorhabditis					
AUTHORS	1 (bases 1 to 31536)					
	Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Cosey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Dublin, R., Favell, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnson, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Lareille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Koopra, A., Saunders, D., Showkneen, R., Smalton, N., Smith, A., Sonnenhammer, E., Staden, R., Stalton, J., Thierly-Mieg, J., Thomas, K., Vaundin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkison-Sproat, J. and Wohlman, P.					
TITLE	2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans					
JOURNAL	Nature	368 (6466), 32-38	(1994)			
MEDLINE	94150718					
REFERENCE	2 (bases 1 to 31536)					
AUTHORS	McMurray, A.					
TITLE	Direct Submission					
JOURNAL	Submitted (25-OCT-1995) Louis, MO 63110, USA. E-mail: jesus@sanger.ac.uk or rwh@nematoe.wustl.edu					
COMMENT	Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.					
	For a graphical representation of this sequence and its analysis see:-					
	http://webcpc.sanger.ac.uk/cgi-bin/display7db-wormaceclass-sequence/subject-E04D5					
	Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.					
	IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.					
	IMPORTANT: This sequence is not the entire insert of clone E04D5.					

FEATURES
source

it may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions..

The true left end of clone E04D5 is at 1 in this sequence. The true right end of clone E04D5 is at 2110 in sequence Z48585.

The true left end of clone ZK673 is at 31437 in this sequence. The true right end of clone T09p3 is at 1412 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence Z49070.

The end of this sequence (31437..31536) overlaps with the start of sequence Z48585.

source

1. .31536
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="II"
/clone="E04D5"
/gene="E04D5.1"
/gene="E04D5.1"
join(6082. .6250,6307. .6379,6432. .6509,6564. .6696, 6750. .6930,7003. .7283,7337. .8179,8228. .8326)
/gene="E04D5.1"
/note="CDNA EST EMBL:D71941 comes from this gene; CDNA EST EMBL:D74691 comes from this gene; CDNA EST EMBL:D76330 comes from this gene; CDNA EST EMBL:D65192 comes from this gene; CDNA EST EMBL:D68540 comes from this gene; CDNA EST EMBL:C11194 comes from this gene; CDNA EST EMBL:C13213 comes from this gene; CDNA EST EMBL:C11881 comes from this gene; CDNA EST yk41207.3 comes from this gene; CDNA EST yk412d7.5 comes from this gene; CDNA EST yk459b6.3 comes from this gene; CDNA EST yk459b6.5 comes from this gene; CDNA EST yk454d9.3 comes from this gene; CDNA EST yk454d9.5 comes from this gene; CDNA EST yk350b11.3 comes from this gene; CDNA EST yk350b11.5 comes from this gene; CDNA EST yk413f5.3 comes from this gene; CDNA EST yk413f5.5 comes from this gene; CDNA EST yk402d12.3 comes from this gene; CDNA EST yk402d12.5 comes from this gene; CDNA EST yk375c1.3 comes from this gene; CDNA EST yk375c1.5 comes from this gene; CDNA EST yk374f8.3 comes from this gene; CDNA EST yk374f8.5 comes from this gene; CDNA EST yk365e8.3 comes from this gene; CDNA EST yk365e8.5 comes from this gene; CDNA EST yk256f12.3 comes from this gene; CDNA EST yk256f12.5 comes from this gene; CDNA EST EMBL:C12803 comes from this gene; CDNA EST EMBL:C12996 comes from this gene; CDNA EST EMBL:C13136 comes from this gene; CDNA EST yk268c5.5 comes from this gene; CDNA EST yk519a6.3 comes from this gene; CDNA EST yk53e1.3 comes from this gene; CDNA EST yk53e6.3 comes from this gene; CDNA EST yk578a6.3 comes from this gene; CDNA EST yk598a11.3 comes from this gene; CDNA EST yk640h2.3 comes from this gene; CDNA EST yk656h4.3 comes from this gene"
/codon_start=1
/protein_id="CAA91279.1"
/db_xref="gi:3875451"
/db_xref="SPTREMBL:Q19052"
/translation="MPAPNTTSPRSDSNFEVTSFSFNNLLPEKFTFSEFRLISOTLII
CINKKKDQYLAAYVBSSEGYLAKRLGKLDADVAEPQNTSDPAKAVAYSNNOCLPFX
DNQVTRFETATRKELICVLEKTRKTLSPKONFLITTEPNAVYCPKRAENKPEP
NVRVYSLADGKHVSTFSPAKESWPEQFSDSESLAARNVGESEFYNNYTPKRDHLL
VEKGATNLSLSPGAPNNAVAYPVAVGSTPARVHVHSESEFVNGRTPEFSKAVM
TANQGSGLLSLASEVDKTNOSYVGEESLTLINIGSESVVPLEKKGSLIYAKAM
NGREAVAYCTEKPAVFYFNRPVNGPIPTDEPNDYFNAVFGVITLREHNGJNAKK
MEFNVAYKKEKLTSTEVNNTTLIPDMADPGQFVYCTTAPLRIDINSYRPHHNGRMIA
ETHSESPKLEWVRWRPWTGYNKAIRKELITDMMAGLPIRKDASHPLANNPACAV
FOAGVATIPPLHAKRLPLGGGSGAPPSAAAPPTGNOONPAOPRANGNSAOPRPOOS
EOERKAFVKEKLEIEIKLVKORVANGQLOPNOMEKLTORENEYLSELSKTLTI"
complement(8619. .11718)
/gene="E04D5.2"
complement(join(8619. .8761,8813. .8988,8985. .10028, 10089. .10283,10330. .10614,10730. .10935,11582. .11629, 11691. .11718))

gene

CDs

gene

CDs

gene
/gene="E04D5.2"
/codon_start=1
/protein_id="CA91282.1"
/db_xref="GI:3673454"
/db_xref="SPTREMBL:Q19055"
/translation="MGFENLYFADIHSITPERCEKIKIRVELSENAQKHLEGFELKYP
INFOLMIVFRRLSPSEIFLATLDCMLICTYLTLEGVAGIYWRKNTLPFLYHR
YIVPEFLATVQVAFIPIILLITPERYLMTEERKRGITPVOVLTLSISPKKASAI
ENERKIRITVIVPCFESVAIRIPVYIAKVTFTPLCDDYFSESIDGTPRATAYEL
YDFHITAIQMFPFVVLNLNLITIKRLVAKERENMPILRGAGTTEVAKASVQV
NLPEPVYLYQAAADYIKESLIHRSSRSLSOLIRNAYIMLAITNSYLNCVHLELI
LERPSPVLYESTDRMOSSTFYILDPSTYICYNASSAIRFIYAKCNPKRQREITDY
IKREKSIETNS"
25834. : 29213
gene
/gene="E04D5.3"
/cds
5010125834..52873,27277..27512,27582..27721,27768..28110,
28158..28410,28491..28611,29092..29213)

gene	29675..31348	/gene="E04D5.4"	
CDS	join(29675..31348)		

```

/db_xref="SPRMBL:Q19054"
/translation="MSINLVHIIILVFSVSHVVTAVIKATGETTRGACQDLDG
VSSSFETTTTASTCADDPNTDCITGYTFLCSNAKYPLDLOGFCGCGSGSTAAP
AACVTMSNO"

```

BASE COUNT	10458 a	5592 c	5636 g	985
ORIGIN				

Query Match	4.7%;	Score 83.8;	DB 34;	Length 31536;
Best Local Similarity	56.1%;	Pred. No. 1.6e-05;		
Matches 179;	Conservative 0;	Mismatches 137;	Indels 3;	Gaps 1;

QY	646	tggaaccaacggttcaacacagttccaatttgcatactatctgcgaacagtttccaataa	705
Db	27783	TGAACCCGGTGATCACTATTAAAGTACGCTTCGAATCGAGATCAAGCTTCTACAAATG	27842
QY	706	gacatgcatcttcgtgaacacgcttgatcacttctgcgcggttgccatcttcgttcga	765
Db	27843	GACTTCGCTGCTCGAAGTTGAGAAAGTGTACTGATGTAAGTTCACTCATGACGGTTA	27902
QY	766	tgaigtgaaggttgatctgtggaattctta---atgctatgatatgtcgtttgataa	822
Db	27903	TGATGGAACAAGGTGAGCACCAGTAACGTATGATATCCATGGATGTTCTGTAGATGG	27962
QY	823	ataattgtctaaataattggaataatccaagatttaatgctgcgcaagaagctcaagt	882
Db	27963	GGTATCTCTTAGATCTTGTAGTACACCTCTGACTTGCACACTGGAAATCGCACCGT	28022
QY	883	ataccaataatcggaatcgatcacagctttctatccaatgccaagatccaglattacat	942
Db	28023	CTTCAAAATTGCTGACAAAGCTGGGCTTTATTTCATTTGTCAAATTCATTAAACATCA	28082
QY	943	agaaccaaatagcgaatgt	961

Db 28083 AGACGTAATTATGGATGT 28101

RESULT 11

LOCUS	CEF53B6	32412 bp	DNA	INV	02-SEP-1999
DEFINITION	Caenorhabditis elegans cosmid F53B6, complete sequence.				

VERSION 281086.1 GI:1627952

SOURCE

REFERENCE
1 (bases 1 to 32412)
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilitida;
Rhabdilitina; Rhabdilitoidea; Rhabdilitidae; Peloderinae; Caenorhabdilis

REFERENCE
AUTHORS

TITLE

JOURNAL
Nature 368 (6466), 32-38 (1994)
0410710

REFERENCE 2 (bases 1 to 32412)

TITLE

JOURNAL

COMMENT

Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail: jesse@amer.ac.uk or rw@nematoide.wstl.edu
Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.
For a graphical representation of this sequence and its analysis see:

http://webc.sanger.ac.uk/cgi-bin/display?db=wormaces&class=Sequence&object=F53b6

Current sequence finishing criteria for the *C. elegans* genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is **NOT** necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

IMPORTANT: This sequence is not the entire insert of clone F53b6. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone F53b6 is at 1 in this sequence. The true right end of clone F53b6 is at 5116 in sequence 281523.

The true left end of clone F32h2 is at 32306 in this sequence. The true right end of clone F08g11 is at 7884 in this sequence. The start of this sequence (1..105) overlaps with the end of sequence 280220.

The end of this sequence (32306..32412) overlaps with the start of sequence 281523.

FEATURES

Source

1.32412

/organism="Caenorhabditis elegans"

```
/db_xref="taxon:6239"
```

```
/chromosome="1
```

```
clone="F53B6"
```

1480. :3649

```
/gene="F53B6.1"
```


CDS	join(1480..1638,1707..1946,2437..2628,2969..3154, 3205..3308,3430..3649) /gene="F53B6.1" /note="predicted using GeneFinder: Similarity to Human platelet-endothelial tetra-span antigen (SMLEPT3_HUMAN)" /codon_start=1 /protein_id="CAB03120.1" /db_xref="GI:3877452" /translation="MKSAGEGRKKKNOILKRDPSKLANPLSEESLYREOPDQFLKN RKINVDQDRLFLSELEYFHSLSINLTLYLALHQPVPFPKSPHNNTLKF OLKMGADSNAYGARGRLIKSYIVTALISLEISICCYIMILARKSOAYELVSP LYDVGRILVILISLILNLTLCFTALFEKRCVTSOAVSIYAVMLITGGTGLN FRDQLTHTPLNLMILSLRELGTGTHDKMLQSNFCKCGVAGDIDMAIMKT VSMYHQRAPKLIPESOCIPSEIERCSNPFQDAPPYVSTCYEPLQNDLHVNM VSMWLCITNAIVQIFVSLIKIRLNLFIKT"
gene	complement(4270..15311) /gene="F53B6.2"
CDS	complement(join(4270..4339,4493..4547,5071..5295, 5368..5662,5891..6148,6204..6752,6818..7287,7546..7676, 8104..8184,12576..12775,13246..13384,13931..14132, 14183..14294,14341..14457,14994..15155,15198..15311)) /gene="F53B6.2"
gene	/note="predicted using GeneFinder: similar to thrombospondin like: cDNA EST EMBL:Z14404 comes from this gene; cDNA EST EMBL:C12571 comes from this gene; cDNA EST yk195c9.3 comes from this gene; cDNA EST yk195c9.5 comes from this gene; cDNA EST yk286g7.3 comes from this gene; yk333e3.5 comes from this gene; cDNA EST yk353h1.3 comes from this gene" /codon_start=1 /protein_id="CAB03121.1" /db_xref="GI:4008366" /db_xref="SPTRMBL:P90884"
CDS	/translation="MKCSYVYVFLIFVLLISFHYDALSNAMSPMSCTCTGGGVS KQLRCLTSCSGSVRRKVCQAKTCEKSLADITCGGEIYSRGCEVGGSRLLT GANFLMRVDDGTPCOAITSRAVCSKSGQIVGCGLISSEFPDAGCGGRTGCDN GKFTIKVSEETACASNCDDIVDMSGARSJASTSOPVYVNAITGHPVEKLCAD LRPVEARPCPMILICPSRMMADWTECVPHCGEETRKREYVCYOTAHNVYHVDTC ENGRPAAEENCYSTSGGRWAGKMSKCTASGGGVRRRHRAVAGSGSDCDGAPROE TTCYAGTICSTATSLDMNDRAVLDGNTFGSMDNHMQARPLVAGEMSTSSPGTC VMSRTVECVANPLISSAPTKIPMSECDDQDEPKLPESCEVRSCLQDESKLSEDEAP OMRTGDMTQCSASLGGOKAKALCIVSTGKSVQSCDARRRPEKSPCNDPCC PFWLTISKSDCSMSGSTARRSVKCAQTVSKTGADAHIVLRDRCHFKKRPQETEC NVACPAWVSLSLRHNKIKLNTKTAQWTECERSCDGERRQWCEIIDSNGKTC RRPVEDCANTRPOTVEVCSFGSRPELLSNRYEONAEQKLTLLGIGVAYLTGCT SIIKCPAKKREKKIKYWKKNKKIKINDAHITKVGANGLRPHARMEDAGYECCTDR LOGVNTLNFKTRDPASRVDLAPKPOIPSTNNRQVOYSKEDVLRBOASVLRHMVSL IEALLTAVNDEKAREQLKRYGNELVARMDIGHMSECRQCTHVAGYQARGLISCKVTEH GEIRNVNSICIESLASVAPETRPCREDCEPRMASOWSECSSQRCVSMIAORRYN TCRTNGTSDVIOHCDITNRPATIMDCNONCKAEMPTSDMGSCSSECTGTGVALRL SCWMISSGRPARNGEOMRRPHSARACVADPELRPPCMPTASALYORASQCDGSRFCD IILFHSDSLLEVOKCCSTCTPYERKKE"
gene	complement(19892..20283) /gene="F53B6.8"
CDS	complement(join(19892..20022,20241..20283)) /gene="F53B6.8"
gene	/codon_start=1 /protein_id="CAB54255.1" /db_xref="GI:5824510"
CDS	/translation="MDFEILLVFLMAVAGTFAGISVSFSSHEPFRHGLLGGGFNRY GGYQSNWYGRK"
gene	complement(20516..21818) /gene="F53B6.4"
CDS	complement(join(20516..20602,20648..20953,20999..21424, 21609..21818)) /gene="F53B6.4"
gene	/note="predicted using GeneFinder: cDNA EST yk193h3.3 comes from this gene; cDNA EST yk193h3.5 comes from this gene; cDNA EST yk411f5.3 comes from this gene; cDNA EST yk411f5.5 comes from this gene" /codon_start=1 /protein_id="CAB03117.1"
gene	/db_xref="GI:3877449"
CDS	/translation="MIRISRCGPFRKSKQSDINDKNOFEVLKEEARCAIARTML VPLAIAYTSLPMVAIAFCAKNKRTYHAKNNKNSKSKSKSSTRGASGSKSR SSKAKHSKSSKSKGSGKSGKSGKSGKSGKSKSKSKSKKRTATTSQSVTVASA TGVSDDKOSNSSKSKSKSKSKSKSKRNRLSDADAKMEKSKSKVALIPTQOOTS OVGSHLAEVNSIKHSKEMNAPAKLOYITGIGNOLELKTSTERRAYKIKGSDNL YRVNRYGFAPRRSSVKIDVLRNGEOKTDLVLLTANANGSTNPHEAFANQAREM MVPVLVA"
gene	23555..24160 /gene="F53B6.5"
CDS	23555..24160 /gene="F53B6.5"
gene	/note="Similarity to Rat Initiation factor 2 associated glycoprotein (SW:12A6_RAT)" /codon_start=1 /protein_id="CAB03119.1" /db_xref="GI:3877451"
CDS	/db_xref="SPTRMBL:P90883"
gene	/translation="MILSDQNFLOTOMKREPQASKNTSEKCEPHGNSNEKPIGSLN GOSIAQCRITHETKTVIVYGGEGMAEMENETIAYETGSGKGFHDMETSHYKNE ELADEKIPRLQSKGLINLIDKNPATLAFRCWIDRLGETRYLMALKKCKYSTASR LMLTLFTDRMMAMVACILIOSPAKTIICKRAULLNLT"
CDS	26258..27459 /gene="F53B6.7"
gene	join(26258..26342,26501..26802,26846..26926,26974..27107, 27153..27288,27334..27459) /gene="F53B6.7"
CDS	/note="predicted using GeneFinder"
gene	/codon_start=1 /protein_id="CAB03118.1"
CDS	/db_xref="GI:3877450"
gene	/translation="MSRESPVSLAQLVERETFNILKRVSSPLVSHASMGILFPNNVR VCYRRGPEHYHDMNMPKLVPRPLHYKEDPDLTEGQIVSEYTGGLNAGIRPE EVPCSPDKSVQTAFLPIKGLSYTINIDPALISROMPTMFOEMLISPRKAFNM GYPINIOYLPESGFIIRAEINEDYMLRDAEFKKNATIEQYVVISDNVVDITRNE HVPNDELQCIKPKTCOMNFTSLKGAQIMSDPLPLTKSLTVLVRPFTVDPLQ SVHPNELNQN"
gene	26276..26348 /gene="F53B6.11"
gene	/note="Tyr Lys K-tRNA: predicted using tRNAscan-SE-1.11; preliminary prediction; similar to tRNA-Lys"
gene	26276..26348 /gene="F53B6.11"
gene	complement(27666..30589) /gene="F53B6.6"
CDS	complement(join(27666..27862,27919..27959,28076..28270, 28905..29001,29330..29417,29467..29606,29648..29915, 30107..30301,30478..30589)) /gene="F53B6.6"
gene	/note="predicted using GeneFinder: similar to cuticulin" /codon_start=1 /protein_id="CAB03116.1"
gene	/db_xref="GI:3877448"
CDS	/translation="MNLSLFTVKTVPYFYVVLLEKVOIHSFKQIVALRFSVOFKNP PILLLILNRIEQLQITILLISTSCSEIIONGVGKREYVCGIDTIRKYNTEHPN GRIYVDGSDKVOHSDADASSPOETRTIGACNMRRORLTHRPISFETMTSR HPFVYGMRAFPISIRCPLESIKGNALIDVGTALPOHVDQEYSPVCAVHLKQGTG HVLRFQVQGVKVTVMRCDDASHYVIGISCYADDSHCKMFKFLVLDRCGSTPPLL POIYERGAISATYNAHYFKYADKYOLFTYVOLCTYKHDSGCEBITPPQSGSHGT
Query Match	4.2%: Score 74.8; DB 34; Length 32412;
Best Local Similarity	59.3%: Pred. No. 0.00056;
Matches 127; Conservative 0; Mismatches 87; Indels 0; Gaps 0;	
QY	393 cacttccattgattcagcaatgttgcgcgtacaacagatcttgatcaacgtgtattt 452
Db	29831 CAATTCCATTGTGAGCATATGCGAGGACCAACATCGATCCACGTGAAATAT 29772
QY	453 ttgtacaacaacgtgtgcatctgattcattcattatatttcaacaagtgtcgtg 512
Db	29771 CATTTTCCCTTCACATGATGACCTCGTTTTCATCTTTTGTGACGGGAATGATCGAG 29712
QY	513 catatcgagtacatgcttttcatatgtaagctgataaaacagttagtgcacagattgag 572

FEATURES		Location/Qualifiers
source		1..152878 /organism="Caenorhabditis elegans" /db_xref="taxon:6239" /clone="Y18D10A"
gene		complement(7920..26256) /gene="Y18D10A.1"
CDS		complement(join(7920..8081,9752..9924,10879..11164,12369..12659,13906..14204,15129..16655,17748..18104,19860..19949,20008..20088,20709..20807,22205..23127,24486..24744,25724..25909,26020..26130,26236..26256)) /gene="Y18D10A.1" /note="predicted using GeneFinder" /codon_start=1 /protein_id="CAA22308.1" /db_xref="GI:3979929"
/translation="MDEDIAEQACSEGCANVRKULGLILMONKRAEMQFSYQA EEVLRMKLVSTDAVRESVQREAEINKEIAEASLKOGYDSEFYKKKASLQEOA DDAEKHQOANVMRKYSVCALTERLANNNDIKSYETVKLLETAGKMTSLDNO VTTLAAKIENSSKTKLDKYLKSVYQVVDALSTSNLSQKSKIKDLVTVRRDVR LEFNSQPMKNLSLSTSDPSISATAEASEDEPHLDVITEPPAPIGEOTLDEDEEE VEDEADDEVDREIROLENEILLQVREKTVRRDPFTAKMIELKAKSRAPRLVETFKV KMPGCIHPTKEDSGLINSTKIHNOPTSDPTLAQTAEKKALLGNOGAASE PESSSIHCKKKTIVRELOTEBMKGESEVKEKARMAEASGRSQAPGAPAPASEL ODPQDFGLMSDPGSDSETEHQOEKHIIPAMVTRRSARLSALVTPKASSSSKMP PPSPSPTPGRGRRRPRLISTMSMEPAAAVTPAPRGRRSRAKVSENTEPLSEA PSAPVRGGRGRSRSTMSITDESEPTSTAAKRSKRAESDEEEDOLKLTANKPEK PKPKTEETETGVDLKKRLRPTAKTTATVTHPPGLRTRKMEMRAPTAATYSSKE KPKNAGSADSSINEEHDEMTILLEQTLDEPOQTSQOEPRISCSSELLDEDFDSE HSGIVSAEELTKNPAPVPEASEASAEPRKIDIEQATPILALALATVSPALAE KRAOENPAVOKISTPTSEISGRAPQALPTSSQPTPGSAAPVDLLSELLSGAKTTTK MEPVOKKISSSEAPPISDSAPTSVHQOTKSPKQLNSKYGDISDEDEEEDEEER MEIVEEEEAPRISDSLOASEPSTATVKEVAVVAVKEFPEIDSTSEAPDEEAP PRAPATKIDNOLADQOASEPPPKARKIPLARKIPKIKSLPAPSSSTDDILMS DILAGAKTTTKTKPAAPTRVAOTTRTKMLAKRRASPTPGATTAPKQOTTKKSID VLPPSSSSSTEPSAPDSASTSSMKKGGAIMEAVPCRPGGKAIKRQOKPIGKME MVQNVKGGKAVNKIKTHLQALDLKIPKEELKPMEEKGKILGDISPLTPEDADVM MEFLRETSADMAVILNRIDANLKLPMNKEENFLQVSVTLNEHDOQLLEFISRI IPELCARESIGSENGRLIRLFCHAVRFESANEPPDVTSPLYOKKSTIRLFOIL LHNPQIKSCSTCYILMSDYAKCRFLVYEMDMKMGSEPHLAFVLMKKSDEQSAV NMLNAKEDVAVYAPPTSNDVTEACINAHKQCIDDDVDVPSLLKSLFLAKTAPAEVL NAVDFLQKMDLIEQFLKPKMEKEPRITAFSVPTVYTKTSEAMLEKRESKQ KNVLRVMDNPPSSSHALFPAALKQISPOVSOLRETIIESEINQYEMTGEDAIRAV ONLVLSIYTFNFPKPTLPPSS"		
gene		complement(26502..29006) /gene="Y18D10A.2"
CDS		complement(join(26502..26573,27346..27405,27490..27580,27624..27788,28732..29006)) /gene="Y18D10A.2" /note="predicted using GeneFinder" /codon_start=1 /protein_id="CAA22309.1" /db_xref="GI:3979930"
/translation="WMNPNKSPPTISINFLPTMSKELLFISAVATASDIVEFGI SNCPTSKRIHQVLVFPYONKGNLSDGLKIDFAVPTGHOVDGYVVRCLHGLEC ALNKLQMSCKHIKQDMLVYTAGCIGQKTAAGCLDPTEEGKIYONCAEEEGEYL LNDENSRYNVAPHSAWLPWIOVNGEBRNNAEFLKDKFSVRLFFEVYLOGLKRSI KN"		
gene		complement(29879..35351) /gene="Y18D10A.3"
CDS		complement(join(29879..30346,31220..31296,31346..31522,34557..34852,35323..35351)) /gene="Y18D10A.3" /codon_start=1 /protein_id="CAA22319.1" /db_xref="GI:3979940"
/translation="MDGSEKAIIGFLQPSLNARFWTKQLLATDEPSESENYOWMR FRISEKRKSVDKLTKEGIEIPARGRNKVGSRVSKMPSISKSMELSLVKLOTPIQ QYMNKGRKSTWHLIISKTVFISGOKLAAQIDEOLEFTKYGKVEQMLDGLAAOA IAAHYKSNVATLCGPGNNGDGFVCARHLQOFGFETPSIYVPKESRNELMKSLVQCE TSSITPTATLPTNIDAPLITVDALFGFSPHPTREPTMLKTVRASGTHVSIIDPS GMDELGAPEGNDVDVIPHVSIVISLPLKLMKMTGPHFLGGRVPSKSLVDEHELLM		

gene		PQYGFEEQIVKLED" 35614..36138 /gene="Y18D10A.4"
CDS		join(35614..35976,36025..36138) /gene="Y18D10A.4" /note="predicted using GeneFinder" /codon_start=1 /protein_id="CAA22310.1" /db_xref="GI:3979931"
gene		/translation="MTGEQTEFAVRSKSNVSEKGGPCRLTNDGRTLSVQSDSS HIDSGTGHFESLDCISLSTTAAVAATATSSSSPKAFQVSYSLFELMLVARY VGVLVTGGMDEKIVVQFPHRLPTRKTIENSAFSSFFCFCLPMITIGFWTLSF" 39072..44375 /gene="Y18D10A.5"
CDS		join(39072..39164,40245..40513,41360..41936,44226..44375) /gene="Y18D10A.5" /note="predicted using GeneFinder; similar to Eukaryotic protein kinase domain; cDNA EST yk30768.5 comes from this gene; cDNA EST yk4519.5 comes from this gene; cDNA EST yk233b4.3 comes from this gene; cDNA EST yk233b4.5 comes from this gene; cDNA EST yk387f2.5 comes from this gene; cDNA EST yk337g9.5 comes from this gene" /codon_start=1 /protein_id="CAA22311.1" /db_xref="GI:3979932"
gene		/translation="MNKQLSCSLKSGKQVTVVASVATDGVQOQVEISYDQKVI GSEGVYFLAKLTSTNEMVAIKKVLQDKRKNRELOIMRKNINVIKLFYFYSGEK KDELVINLILEEVPETVYVARYVSKORQIDPMIYKVLQMLLSLAIHSIGR DIRQNLILDEPSVYLKCDRESAKYLVANERNVYICRTRYRABELTGATNYTNSI DWASQVTAWEILLQOPIFGDSGDQVLEIILKVGTPREIOISNPNYKCFKFPQI KAPNKNKFRVHTPPAEALDISKILIEYTPSTPQAOACQAHFPELRLNPARLPBSGR PLPTLEMDGPMGTGEVSTSGVAPSA" 47525..53224 /gene="Y18D10A.6"
CDS		join(47525..47593,48112..48208,48610..48751,49616..50045,50501..50818,51535..51971,52106..52364,52847..53224) /gene="Y18D10A.6" /note="similar to sodium/hydrogen exchanger family" /codon_start=1 /protein_id="CAA22320.1" /db_xref="GI:3979941"
/translation="MNLSEBOVLARNNIELMFTNSLKLANGVTQWLWESLNOGTA TSGILMFAKLTALISIFLLIPQYATDSDSASASVSGAVKSEDTVAANKVIDP GEAIDANATSLBOHGAALVGNVSEKRSALFELFVLMTLVLMVILVSKIHMP ESLAIYALGALIGSILSYSRDMSETEALSPVEFLVYLPFIFENAYLNLNGYESN EYVLTIFAEFTGTSAMVAGALYILGALIEGTEFFECFPAAMISYADVGTALAI ROAVYESLLMYLVGESLNDNVAISYLAATPLNRKAEFNSLPASETTSAPVYTE MFFESACLGIGLISALIFKHVNDLKPTLSLEFALLIFSYIPIGAEALDLSGTMAI LFCGISMSQFTRHNVSPIAQITFRHRTFRTISFVAETSTFAYIGMAFFTILKLFAPWLI FWSVYCLIGRACNVFLAYLVNOCGRQOISMKNOIIMPFSGMRGAVCFALVLYMDL DKKRSILTLVYLFILFTTIFLGSALEFISFINCYENRQKRRRTPRNKESTGN SSALMKSTQOEMSFGSDMGPKRSALDNTSASGRMRBLFVRKFAITRELRNPKLA ALTKRAASDQMTDSDPDEFGGCGGAVGGGRKMDVTPTRGSGSGNSSDVIYSAGCG GVSGEHLHLSGSDSTNEF" 54375..62943 /gene="Y18D10A.7"		
gene		join(54375..54405,54825..54924,54984..55168,55357..55538,56162..56529,56806..57056,57123..57193,57641..57905,58630..58895,58941..59066,59124..59210,59257..59373,59630..59775,61492..61610,61724..61783,61846..61959,62188..62409,62843..62943) /gene="Y18D10A.7" /note="predicted using GeneFinder" /codon_start=1 /protein_id="CAA22312.1" /db_xref="GI:3979933"
/translation="MSQNHILLPDSFKNSMTISKITRAAPLEQPMANIVAKYCLEFVAK YMPPTIILITICLSMGIILNFKIVRVNVLVPLNATWTEAEVGEENAKDDDH FYPGKDILROGIVLIVNACDGNVLRQOADDPLKIIDMLINVLKLSAGRIPTYKQ VCLHONDECSFPAKLLANITYSKHONSMENTIYTPYASATATPDIOSKYLGVSL DYDGHVENASAMLILYQLKNEKWLSDRDEDLAKIQSGEAPSELNLYYHSATFD		

Query Match

3.9%; Score 68.8; DB 34; Length 152878;


```
/note="predicted using GeneFinder; preliminary prediction"  
/codon_start=1  
/protein_id="CAB60892.1"  
/db_xref="GI:6434393"  
/translation="MPNEIEEGSGLDVARSRLAGKKWYCSILLVNLINMRYTIV  
GWSRLATFDDIDDKGQGLQTVFIVFVFAPLFGYLDGRYNRKMIMTIGICIMILA  
VASSFCGEGHYLFLCRGIVIGEASYSIAPVYLSDFSGGLSRVLMFYPFAP  
VSGSGFISGSSISOATDSMWGVRFSPYIGIACGLMLMLDEPVRGACDARONGD  
EADLIGDIKLYMSITFEYLAASAISAFPSIGTMSWMPQYGFSAVYHNPKVET  
ELTOINLJEGITTCMAIGTAVTSGTILSRMRDGSIIFFNHATEKADVYICALSMPVA  
LPLFPATFIATYSTNGCLILYFPAIMSKLMSVNVYDLYMYVANRRATLAVQTM  
VAHLFGDASPIYIGVLSMDLGGDDASAVGHRFALQKALVPTFMDLVAGATFLATF  
FVEDRKEALVOMDENFWYTELKADAMSESGRGSLLGRVLSDFLPPDIETLASV  
DEDDGINLEEDACTIKFDSIRRRNDNSYENGQNAADLSDETDEWDVQRPDSGDDDP"  
BASE COUNT      24962 a 12141 c 12530 g 23596 t  
ORIGIN
```

```
Query Match      3.8%; Score 67.8; DB 34; Length 73229;  
Best Local Similarity 52.3%; Pred. No. 0.0079;  
Matches 150; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
```

```
QY 1493 ttatagcttcgaaaagcttatcatcattcagaatcctttatatgatactatgt 1552  
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 51833 TTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTC 51892  
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 1553 aaatgttcacatcagccacgaatagttcgttctgtaatacatalcatcaacttg 1612  
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 51893 TCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTT 51952  
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 1613 tcctattttcctaagaattacatttctgataatacacaaattataccttgatlg 1672  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 51953 TTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTC 52012  
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 1673 cccaatttataggacatcattccatactctgtaacaactcactatctgcatltgc 1732  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 52013 TCTCATTTTTCCTTCATTTTTCCTTTTGTGTTTACATTTTTCATTTTTCATTTTTC 52072  
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 1733 aaltaaaaagatltcatltgtgaaaaaataaaaaaataaaaaa 1779  
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 52073 TTTTFAAAATCATTTTATGACAAAAAAATGACAAAAAA 52119  
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: April 16, 2000, 04:42:23
Job time: 10352 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 16, 2000, 01:58:21 ; Search time 265.24 Seconds
(without alignments)
1678.072 Million cell updates/sec

Title: US-09-323-427-1
Perfect score: 1779
Sequence: 1 gttcaattaccacaagttyg.....aaaaaaaaaaaaaaaaaaaaa 1779

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues
Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	67	3.8	605	T31530	Human 3' apolipoprotein B
2	58.2	3.3	6152	T78867	P. falciparum live
3	55.4	3.1	19124	T72882	Plasmodium var-7 g
4	55.2	3.1	6152	T78867	P. falciparum live
5	54	3.0	5852	Q11710	Dictyostelium plas
6	54	3.0	8920	Q62924	Carbamoyl-phosphat
7	51.8	2.9	1052	N90224	Malaria-specific D
8	51.8	2.9	9789	T41852	CDNA encoding Plas
9	51	2.9	605	T31530	Human 3' apolipoprotein B
10	51	2.9	19124	T72882	Plasmodium var-7 g
11	50.8	2.9	110000	X20248_02	Continuation (3 of
12	50.6	2.8	4590	N60472	Sequence encoding
13	49.8	2.8	4590	N60472	Sequence encoding
14	49.2	2.8	715	V68977	DNA molecule encod
15	49.2	2.8	6124	Q03568	Sequence encoding
16	48.6	2.7	1052	N90224	Malaria-specific D
17	48.6	2.7	1052	N90224	Malaria-specific D
18	48.6	2.7	8920	Q62924	Sequence encoding
19	47.6	2.7	3101	Q02047	Carbamoyl-phosphat
20	47.4	2.7	4673	Q27189	Sequence encoding
21	47	2.6	2504	T62359	P. yoelii SSP2 ant
22	47	2.6	2504	T62359	Schistosaccharomyce
23	46.8	2.6	1942	T73286	S. pombe origin of
24	46.8	2.6	110000	Q12540	GS3A promoter elem
25	46.6	2.6	1711	V33136	Continuation (10 o
26	46.4	2.6	9636	Q67190	Plasmodium berghai
27	46.2	2.6	5181	Q80911	P. falciparum tran
28	46.2	2.6	110000	V21209_04	Plasmodium falcipa
29	46.2	2.6	110000	V21209_05	Continuation (5 of
30	45.6	2.6	1939	Q34622	Continuation (6 of
31	45.6	2.6	6644	X33181	Cytosolic GS3A gln
32	45.6	2.6	7372	X33181	Base sequence of t
33	45.6	2.6	7797	X33180	Base sequence of t
34	45.6	2.6	7996	X33184	Cowpox virus bsr f

35	45.6	2.6	110000	X20248_04	Continuation (5 of
36	45.6	2.6	116277	X20249	Borrelia burgdorferi
37	45.4	2.6	2338	Q54656	Plasmodium PNXL compo
38	45.4	2.6	110000	T58840_4	Continuation (5 of
39	45.2	2.5	2430	V22750	Continuation (5 of
40	45.2	2.5	2430	V22750	Babesia microti BM
41	44.8	2.5	2503	Q53480	Babesia microti BM
42	44.8	2.5	4053	X52266	PNPX30 xylanase CD
43	44.6	2.5	144	Q12515	Protein PRO326 CDN
44	44.6	2.5	2104	Q25273	CSP-2 peptide from
45	44.6	2.5	5852	Q11710	Sequence encoding
					Dictyostelium plas

ALIGNMENTS

RESULT	ID	Score	Length	DB ID	Description
1	T31530	3.88	605	T31530	Human 3' apolipoprotein B
2	T31530	3.3	6152	T78867	P. falciparum live
3	T31530	3.1	19124	T72882	Plasmodium var-7 g
4	T31530	3.1	6152	T78867	P. falciparum live
5	T31530	3.0	5852	Q11710	Dictyostelium plas
6	T31530	3.0	8920	Q62924	Carbamoyl-phosphat
7	T31530	2.9	1052	N90224	Malaria-specific D
8	T31530	2.9	9789	T41852	CDNA encoding Plas
9	T31530	2.9	605	T31530	Human 3' apolipoprotein B
10	T31530	2.9	19124	T72882	Plasmodium var-7 g
11	T31530	2.9	110000	X20248_02	Continuation (3 of
12	T31530	2.8	4590	N60472	Sequence encoding
13	T31530	2.8	4590	N60472	Sequence encoding
14	T31530	2.8	715	V68977	DNA molecule encod
15	T31530	2.8	6124	Q03568	Sequence encoding
16	T31530	2.7	1052	N90224	Malaria-specific D
17	T31530	2.7	1052	N90224	Malaria-specific D
18	T31530	2.7	8920	Q62924	Sequence encoding
19	T31530	2.7	3101	Q02047	Carbamoyl-phosphat
20	T31530	2.7	4673	Q27189	Sequence encoding
21	T31530	2.6	2504	T62359	P. yoelii SSP2 ant
22	T31530	2.6	2504	T62359	Schistosaccharomyce
23	T31530	2.6	1942	T73286	S. pombe origin of
24	T31530	2.6	110000	Q12540	GS3A promoter elem
25	T31530	2.6	1711	V33136	Continuation (10 o
26	T31530	2.6	9636	Q67190	Plasmodium berghai
27	T31530	2.6	5181	Q80911	P. falciparum tran
28	T31530	2.6	110000	V21209_04	Plasmodium falcipa
29	T31530	2.6	110000	V21209_05	Continuation (5 of
30	T31530	2.6	1939	Q34622	Continuation (6 of
31	T31530	2.6	6644	X33181	Cytosolic GS3A gln
32	T31530	2.6	7372	X33181	Base sequence of t
33	T31530	2.6	7797	X33180	Base sequence of t
34	T31530	2.6	7996	X33184	Cowpox virus bsr f

Db 314 ATTACATATTTTATTAATAAAGTTTATTAATTCATATTTTATTAATAAAGTTTANA 373
 Qy 1670 ttgcccaattttatcgccatcttcctatctctgtaaacatctcatttcgcatat 1729
 Db 374 ATTACATATTTTATTAATAAAGTTTATTAATTCATATTTTATTAATAAATTA 433
 Qy 1730 tgcacatlaaaagatcttcatttgcgaaaaaataaaaaaataaaaaa 1778
 Db 434 AAGCTATTTTATTAATTCATATTTTATTAATAAAGTATTTATTAATTCACATA 482
 RESULT 2
 T78867
 ID T78867 standard; DNA: 6152 BP.
 AC T78867;
 DE 08-OCT-1997 (first entry)
 KW Plasmidium falciiparum; pre-erythrocyte; liver stage antigen; serum;
 KW glycosyl-phosphatidylinositol membrane anchoring sequence; antibody;
 KW vaccine; immunotherapy; malaria; ds.
 OS Plasmidium falciiparum.
 FH Key Location/Qualifiers
 FT cds 77..5605
 FT /tag= a
 FT /product= liver stage antigen-3
 FT exon 77..274
 FT /tag= b
 FT /number= 1
 FT intron 275..442
 FT /tag= c
 FT /number= 1
 FT exon 443..5602
 FT /tag= d
 FT /number= 2
 FN W09641877-A2.
 PD 27-DEC-1996.
 PR 12-JUN-1996; F00894.
 PR 13-JUN-1995; FR-007007.
 PA (INSP) INST PASTEUR.
 PI Daubersies P. Druilhé P;
 DR WPI: 97-065464/06.
 PT Plasmidium falciiparum poly:peptide(s) and related nucleic acids -
 PT derived from the liver stage antigen-3, useful for malaria vaccine
 PT prodn. and diagnosis
 PS Claim 20: Fig 1A-E: 69pp: French.
 CC This sequence corresponds to the genomic sequence encoding a Plasmidium
 CC falciiparum strain K1 pre-erythrocytic liver stage antigen-3 (LSA-3)
 CC protein (W24790). The gene sequence was isolated by screening a
 CC P. falciiparum strain T9/96 library with the serum from a missionary
 CC treated by prophylaxis (for strain T6/96 see FR9101286). Of 20 clones
 CC isolated, clone 7295 was used to screen a library generated from Thai
 CC strain K1. One clone contained a 6.85 kb insert including the sequence
 CC presented here. The gene organisation comprises a first exon, a short
 CC 168 bp intron and a 5 kb second exon containing a 1.8 kb region encoding
 CC 7 blocks of 4 amino acid repeats and a 3' hydrophobic region
 CC corresponding to a glycosyl-phosphatidylinositol membrane anchoring
 CC sequence. The invention relates to new polypeptides of at least 10
 CC amino acids derived from the LSA-3 polypeptide with the exception of
 CC peptides W24791-4. The LSA-3 peptides can be used to raise antibodies
 CC and as vaccines for immunotherapy of malaria.
 CC Sequence 6152 BP; 2725 A; 463 C; 1215 G; 1749 T;
 SQ
 Query Match 3 3%; Score 58.2; DB 1; Length 6152;
 Best Local Similarity 47.9%; Pred. No. 0.003;
 Matches 197; Conservative 0; Mismatches 213; Indels 1; Gaps 1;
 Qy 1370 atatatcgtcttctatcatccctctaatcaatttagtacaacataata 1429
 Db 5728 ATATATATGTAATCTTTTACAAATTTTAAATTTTAAATTTATATATTAATTTTA 5787
 Qy 1430 gtaatgagaataaataactgtaatacaataagatgataatttcacataaactctctat 1489

Db 5788 TATTTTTCATATTAATTAATTTTATTTTCAATATTTTATTTTATTAATAATGTTTTTAC 5847
 Qy 1490 cgcctttatagcttctgaaagcctatcattatcgaacatctttatgatcactat 1549
 Db 5848 AGAGTTTATGTTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5907
 Qy 1550 tgaatcgttcatcatcattgagccatgaatagttcgttctatcatcattcacaac 1609
 Db 5908 GATATATGTAATTAATTAATTTTGTGTTTATTAATAATTTATTAATTAATTAATTAATTA 5967
 Qy 1610 ttgtccatcttat-tctaacagttatcatttgcgataatacaacataaccctgt 1668
 Db 5968 ATATATATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6027
 Qy 1669 attgcccaattttatcgccatcttcctatctcgttaaacatcattcatttcgacat 1728
 Db 6028 TAAGCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 6087
 Qy 1729 tgcacatlaaaagatcttcatttgcgaaaaaataaaaaaataaaaaa 1779
 Db 6088 TTAATATTTTATTAATAAATTTTTCCTAAATAAATAAATAAATAAATAAATAAATAA 6138
 RESULT 3
 T72882
 ID T72882 standard; cDNA: 19124 BP.
 AC T72882;
 DE 12-SEP-1997 (first entry)
 KW Plasmidium var-7 gene.
 KW DBL gene family; SAMP; sialic acid binding protein; vaccine; therapy;
 KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 KW Plasmidium var-7.
 OS Plasmidium var-7.
 FH Key Location/Qualifiers
 FT exon 7317..15139
 FT /tag= a
 FT /number= 1
 FT intron 15140..16205
 FT /tag= b
 FT /number= 1
 FT exon 16206..17552
 FT /tag= c
 FT /number= 2
 FT /note= "no stop codon given"
 FN W09640766-A2.
 PD 19-DEC-1996.
 PR 07-JUN-1996; U09508.
 PR 07-JUN-1995; US-487826.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
 PI Wellens TE;
 DR WPI: 97-052231/05.
 DR P-PSDB: W2475.
 PT New malaria vaccines - contains cysteine-rich DBL family protein
 PT binding domains homologous domains of the Duffy and sialic acid
 PS binding proteins
 PS Claim 4: Page 56-61, 96pp: English.
 CC This sequence represents the var-7 gene of Plasmidium. Var-7 belongs to
 CC the Duffy binding like (DBL) family of genes which have homology to the
 CC Duffy antigen binding protein (DABP) and sialic acid binding protein
 CC (SABP) conserved regions (see T72889 and T72888 respectively). The var
 CC family of genes modulate cytoadherence and antigenic variation of
 CC Plasmidium infected erythrocytes. SAMP and the Duffy antigen binding
 CC supernatant after infected erythrocytes release merozoites. DABP and SAMP
 CC mediate the binding of merozoites and schizonts to the erythrocyte
 CC surface. These proteins are necessary for erythrocyte invasion by the
 CC parasite. This sequence can be used in the compositions of the invention.
 CC The compositions are for the treatment and prevention of malaria, and
 CC comprise either a nucleotide sequence or encoded polypeptide of the

PF 02-NOV-1990; AU0530.
 PR 02-NOV-1989; AU-007187.
 PA (UYMA-) MACQUARIE UNIT.
 PI Slade MB, Chang ACM, Williams KL;
 DR WPI: 91-164194/22.
 P-PSDB: R11988.
 PT Polypeptide facilitating extra-chromosomal replication - of
 recombinant plasmid in Dictyostelium species
 PS Claim 15; Fig 1; 90pp; English.
 CC The sequence of pdp2 has been found to contain the putative open
 CC reading frame indicated in the Features Table. The possible ORF is
 CC flanked by regions with similarity to promoter and polyadenylation
 CC signals of known Dictyostelium genes. The RNA and polypeptide
 CC product of the Rep gene have not, however, been detected. It is
 CC believed that the product is produced in low amounts to positively
 CC regulate initiation of plasmid replication. The polypeptide may also
 CC contain regions that act as negative regulators of plasmid copy
 CC number. See also Q11711 and Q11712.
 SQ Sequence 5852 BP; 2298 A; 651 C; 708 G; 2195 T;

Query Match 3.0%; Score 54; DB 1; Length 5852;
 Best Local Similarity 48.0%; Pred. No. 0.021;
 Matches 218; Conservative 0; Mismatches 230; Indels 6; Gaps 2;

QY 1332 aataatgtagaatacgcgaagaataaactgcatatataatcgttcttctat 1391
 DB 1583 AAAAATTATTAACCAACCTAGTTATATCTTCCCTCTTTTCTTTT 1642
 QY 1392 catccttctaataactatttagcacaacataatagcaggaataatcctgc 1451
 DB 1643 TGTCAAGACCTTTTCTTTTGTGATGACACTTTTAAAAAATAAAT 1702
 QY 1452 aatacaataagtagatcttccatcaaaactcttcacgctttagcttgaagaag 1511
 DB 1703 GTTAAATATCTATTTGATGATCATTTTCTAGTTTCTTTTGAATGATATAAAA 1762
 QY 1512 ctatcatcatatcaglaactcttataatgacatactatgaaatgcttcatcaggc 1571
 DB 1763 TAAATGGCATGATGATATTAATTAATTAATTAATTAATTAATTAAT 1822
 QY 1572 catgaatagttcgttctgataatcatcatcctcaactgctcctattatctcaag 1631
 DB 1823 AATTTCTACTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1882
 QY 1632 ttatcatctgtgataataacac---aaatataccctgtatgcccaatttata 1687
 DB 1883 TACATCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1942
 QY 1688 c--atcaattcctatctcgttaaacatcactatctgcatatgcaattaaagat 1745
 DB 1943 ATTAATAACTTAATGATGATTAACCACTTTTCTTTCTTTTCTTTTCTTTT 2002
 QY 1746 ttcatctgcaaaaaaataaaaaaataaaaaa 1779
 DB 2003 TTTACTTTGAAAAAATAAAAAAATAAAAAA 2036

RESULT 6
 ID 062924 standard; cDNA; 8920 BP.
 AC 062924;
 DT 06-DEC-1994 (first entry)
 DE Cardamoyl-phosphate-synthetase II.
 KW Cardamoyl-phosphate-synthetase II; CPSII; pscpsII gene;
 KM malaria; ss.
 OS Plasmodium falciparum.
 FH Location/Qualifiers
 FT cds 1226..8401
 FT /tag= a
 FT /EC_number= 6.3.5.5
 PD MO9412643-A.
 PD 09-JUN-1994.

PF 02-DEC-1993; AU0617.
 PR 03-DEC-1992; AU-006206.
 PR 16-DEC-1992; AU-006380.
 PA (UNTX) UNISEARCH LTD.
 PI Flores MV, Osullivan WJ, Stewart TS;
 DR WPI: 94-200271/24.
 P-PSDB: R55694.
 PT Nucleic acid encoding cardamoyl phosphate synthetase II -
 PT isolated from Plasmodium falciparum, used to develop prods. for
 PS the treatment of malaria.
 PS Disclosure: Page 6-16; 31pp; English.
 CC The cDNA sequence encoding the cardamoyl-phosphate-transferase II
 CC (CPSII) of Plasmodium falciparum was determined. The cDNA encodes
 CC a protein that includes 2 insert sequences not found in other CPSII
 CC proteins. The first separates the putative structural subdomain and
 CC the glutamine subdomain of the glutamine- amidotransferase subunit
 CC of CPSII, while the second separates 2 ATP binding subdomains of the
 CC CPSII subunit, CPSa and CPSb.
 SQ Sequence 8920 BP; 3836 A; 774 C; 1232 G; 3078 T;

Query Match 3.0%; Score 54; DB 1; Length 8920;
 Best Local Similarity 47.6%; Pred. No. 0.022;
 Matches 195; Conservative 0; Mismatches 210; Indels 5; Gaps 1;

QY 1370 atataatcgttcttctatcatccttcaataactaatttagctacaatatata 1429
 DB 245 AAAAATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 304
 QY 1430 gtagtaggaataaactactgtaataacaaatgtagatlttcatcaaaccttctat 1489
 DB 305 AT 364
 QY 1490 cgttttatagctctcgaagaactatctatcatcgaactttatcatcactat 1549
 DB 365 TTTATTAATTTACTTGAATATTAACCTTTATATATATTTCCAAATTAATGATACAT 424
 QY 1550 tctaaatgttcatcatccttagcgaatagatctgcttctgataatcactcaac 1609
 DB 425 TATTAATATTTGATGTGATACATTAATATAGTT-----TTACACTTCTTAAATATA 479
 QY 1610 tttcctatcttattcctaacagtttcatctgtgataatcactcaactctgta 1669
 DB 480 CCATCTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 539
 QY 1670 ttgccaattttatggagatcatctccatctcgttaaacatcactatctgcatat 1729
 DB 540 ATTTATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 599
 QY 1730 tgcattaaagaatattcattctgtgaaaaaataaaaaaataaaaaa 1779
 DB 600 AGTAATTTCTACTTAATTTTAAAAAATAAAAAAATAAAAAAATAAAAAA 649

RESULT 7
 ID N90224 standard; DNA; 1052 BP.
 AC N90224;
 DT 1-NOV-1989 (first entry)
 DE Malaria-specific DNA insert of clone 41-2.
 KW DNA; malaria; clone 41-2; Plasmodium falciparum; protein; vaccine.
 OS Plasmodium falciparum
 PN EP-322712-A.
 PD 05-JUL-1989.
 PR 20-DEC-1988; 121299.
 PR 30-DEC-1987; DE-831351.
 PA (BERM) Behringwerke.
 PI Knapp B, Hundt E, Enders B, Kupper H;
 DR WPI: 89-194071/27.
 DR P-PSDB: P90417.
 PT New antigenic proteins from Plasmodium falciparum - new encoding nucleic
 PT acid sequences and derived antibodies, useful in vaccines, diagnosis etc.
 PS Claim 1; Table 17; 25pp; German.


```

Db      8139  TTTATTTTAAATATATATTTCTTTTAAACATTTTATTTTATTTATGATATATATATA 8080
Qy      1588  tgtatattatcatattcaactgttcctatttattcacaagttatcatattgtgata 1647
Db      8079  TATATTTTAAACATTTTCTTTCTTTTATTTTATTTAATAATTTTATTTTATTTTA 8020
Qy      1648  ataccacaattacacctgattgtgcccaattttatgagccatcttcctattctgtaa 1707
Db      8019  TTTCTATATTAATTTTATTTTATTAAT-CATTTATTTTATTAATATATATTTTATTTTAA 7961
Qy      1708  acaattcattatcgtcatcgtcaattaaagttatcattcattgtgaaaaa 1762
Db      7960  ATTTTCATTTCTTTTATTTTATGAAATTTTATTTTATTTTATTTTATTTAATAA 7906

RESULT  9
T31530/c
ID      T31530 standard; cDNA; 605 BP.
AC      T31530;
DT      15-SEP-1996 (first entry)
DE      Human 3' apolipoprotein B SAR element clone Rh32.
KW      Erythropoietin; EPO; anaemia; gene therapy; vector;
KW      scaffold attachment region; SAR element; apolipoprotein B;
KW      transgenic animal; ss.
OS      Homo sapiens.
PN      WO9619573-A1.
PD      27-JUN-1996.
PF      18-DEC-1995; CA0696.
PR      19-DEC-1994; US-358918.
PA      (CANG-) CANGENE CORP.
PI      Delcive G;
DR      MPI; 96-309587/31.
PT      Recombinant DNA molecule expressing mammalian erythropoietin
PT      useful to transform cell lines, and for gene therapy, e.g. of
PT      anaemia and other red blood cell disorders
PS      Claim 7; Page 59-60; 84pp; English.
CC      Human apolipoprotein B (apoB) scaffold attachment region (SAR)
CC      element clones Rh32 (T31530) and Rh10 (T31531) respectively carry
CC      the 3' human apoB SAR element and the distal 1212 bp 5' human apoB
CC      SAR element and 1317 bp proximal sequence. These SAR elements
CC      co-map with the boundaries of the human apoB gene chromatin
CC      domain. A novel recombinant DNA molecule adapted for transfection
CC      of a host cell comprises an erythropoietin (EPO) cDNA (T31529) or
CC      genomic clone (T31532) operably linked to an expression control
CC      sequence and to the 5' and 3' SAR elements. The SAR elements
CC      increase expression of the recombinant EPO in stable, long-term
CC      mammalian cell cultures.
SQ      Sequence 605 BP; 278 A; 18 C; 35 G; 274 T;

Query Match      2.9%; Score 51; DB 1; Length 605;
Best Local Similarity 48.9%; Pred. No. 0.052;
Matches 195; Conservative 0; Mismatches 200; Indels 4; Gaps 2;
Qy      1328  taaaataatgttagaatcatcgagcaataataaactgcga--tatatatcgtttct 1385
Db      403  TATTAATATATGTAATTAATAAATATGTAATTAATAACATTTTATTAATAAATATGTAAT 344
Qy      1386  tcttatcatcctctctaactaattttagctacaataataatagatagtgagaataat 1445
Db      343  TATTAACATTTTATTAATAAATATGTAATTAATAACATTTTATTAATAAATATGTAAT 284
Qy      1446  tactgtaatacaataagtgatatttcatcaaaactctctcgttcgtttatagcttct 1505
Db      283  TATTAACATTTTATTAATAAATATTAATTAATAATATTTTAATTAATAAATATTTAAT 224
Qy      1506  gaaaagcttatcatattcatcgaatcctttatatgatactatcgtt--aaatgttcat 1563
Db      223  TATTAATATTTTATTAATAAATATTAATTAATAATATTTTATTAATAAATATTTAAT 164
Qy      1564  cattagagcattgaatggttcgttcttatatacatcattacaactgtccatttat 1623

```

```

Db      163  TATTAATATTTTATTAATAAATATTTAATTAATAATATTTTATTAATAAATATTTAAT 104
Qy      1624  tctaaagatttaccatttctgtgataatacaaatataactgtattgtcccaatttca 1683
Db      103  TATTAATATTTTATTAATAAATATTTAATTAATAATATTTTATTAATTAATAAATATTTAAT 44
Qy      1684  tgggacatcattcctattcctgtaacaatcattcattt 1722
Db      43  TATTAATATTTTATTAATAAACAATTAACCTCATCT 5

RESULT  10
T72882/c
ID      T72882 standard; cDNA; 19124 BP.
AC      T72882;
DT      12-SEP-1997 (first entry)
DE      Plasmodium var-7 gene.
KW      DBL gene family; SABB; sialic acid binding protein; vaccine; therapy;
KW      Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
KW      DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
KW      Plasmodium; ss.
OS      Plasmodium vivax.
PN      Plasmodium falciparum.
PT      Key
PT      Location/Qualifiers
FT      exon
FT      7317..15139
FT      /tag= a
FT      /number= 1
FT      Intron
FT      15140..16205
FT      /tag= b
FT      /number= 1
FT      Exon
FT      16206..17552
FT      /tag= c
FT      /number= 2
FT      /note= "no stop codon given"

WO9640766-A2.
PD      19-DEC-1996.
PF      07-JUN-1996; U09508.
PR      07-JUN-1995; US-487826.
PA      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI      Chlunis C, Miller LH, Peterson DS, Sim KL, Su X;
PI      Wellens TE;
DR      WPI; 97-052231/05.
P-PSDB; W22475.
PT      New malaria vaccines - contains cysteine-rich DBL family protein
PT      binding domains homologous domains of the Duffy and sialic acid
PT      binding proteins
PS      Claim 4; Page 56-61; 96pp; English.
CC      This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to
CC      the Duffy binding like (DBL) family of genes which have homology to the
CC      Duffy antigen binding protein (DABP) and sialic acid binding protein
CC      (SABP) conserved regions (see T72889 and T72888 respectively). The var
CC      family of genes modulate cytoadherence and antigenic variation of
CC      Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC      protein (DABP) are soluble proteins that appear in the culture
CC      supernatant after infected erythrocytes release merozoites. DABP and SABP
CC      mediate the binding of merozoites and schizonts to the erythrocyte
CC      surface. These proteins are necessary for erythrocyte invasion by the
CC      parasite. This sequence can be used in the compositions of the invention.
CC      The compositions are for the treatment and prevention of malaria, and
CC      comprise either a nucleotide sequence or encoded polypeptide of the
CC      var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
CC      genes having homology with conserved regions of DABP and SABP. The
CC      compositions are used for the treatment and prevention of malaria. They
CC      are also used in the preparation of vaccines for inducing a protective
CC      immune response in a mammal to Plasmodium merozoites (especially
CC      Plasmodium falciparum or Plasmodium vivax).
SQ      Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T;

Query Match      2.9%; Score 51; DB 1; Length 19124;
Best Local Similarity 45.6%; Pred. No. 0.1;
Matches 180; Conservative 0; Mismatches 215; Indels 0; Gaps 0;

```

```
OY 1328 taataaatgttagaatacgaagaacaaataaaactgcacatataatcgttcttc 1387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 15944 TTAATTAATATTTTATTTTATTTTATTTTCAATTAATTTTATTTTATTTTATTTA 15885
OY 1388 ttatcactcttaactaaattttagtcaacaatatatagttgtaggaataatta 1447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 15884 ATTTAAATTTTATTTATTTATTTTATTTTATTTTAAATTTTATTTTATTTATTTT 15825
OY 1448 ctgtaacaataagtgatatttcatccaactcttcacgcctttatagcttcga 1507
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 15824 TTTTATTTTAAATTAATTTTATTTTATTTTATTTATTTTATTTTATTTTAACTTTT 15765
OY 1508 aaagcttaccatctacgaatacctttatagacactatlgtaaatggttccalc 1567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 15764 ATTTTATTTTATTTATGATATATATTTTATTTTAAATTTTATTTTCTTTTATTTT 15705
OY 1568 aggcataaataagttcgttggttatcatcactatatacactggttccatttctca 1627
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 15704 TTTTATGATATATATTTTATTTTATTTTAAATGTTTATTTTCTTTGCTTTTATTT 15645
OY 1628 acagttatcatctgtgataataatcaaatattacactgttatgccaattcttaag 1687
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 15644 TTTTATTAATATTCATTTTATTTTATTAATAAATTTTATTTTATTTTATTTTGAATAATCTT 15585
OY 1688 catcattctcattctctgtaaacacattccattat 1722
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 15584 TTTCAATTTTATTTCTATCAAAATTTATATTTTAT 15550
```

```
RESULT 11
X20248_02/c
Continuation (3 of 10) of X20248 from base 200001 (Borrelia burgdorferi polynucleotide s
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248
WP Fragment Name Begin End
WP X20248_00 1 110000
WP X20248_01 100001 210000
WP X20248_02 200001 310000
WP X20248_03 300001 410000
WP X20248_04 400001 510000
WP X20248_05 500001 610000
WP X20248_06 600001 710000
WP X20248_07 700001 810000
WP X20248_08 800001 910000
WP X20248_09 900001 910715
```

```
Query Match 2.9%; Score 50.8; DB 1; Length 110000;
Best Local Similarity 44.8%; Pred. No. 0.16;
Matches 242; Conservative 0; Mismatches 292; Indels 6; Gaps 1;
```

```
OY 1221 tctgcatgacaccattcggcttcctcaatggttcaagtttaagcaattgcatgattgctg 1280
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 41025 TTTTCAATTTGAAATTTTATTTTCTATCTTTTGTATGCTTCAAGTATTTGTTATTTGTTT 40966
OY 1281 ccgctaatattacacattcgtttaaatctgcaccaatcagaagcataaaaaataatglt 1340
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 40965 TTGATGATATTAATTTCTGGGCTTAAGTTGCTTGAAGAAGACAGATAAATTAATTTT 40906
OY 1341 agaatcacgaagcaatatataaactgcacatatatctggttcttcttatacctctct 1400
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 40905 CTGAGTGTGCAAAATTAATTAATAAAGACCGCTTATTTCTCTTTTCAATTTCTCAT 40846
OY 1401 aataactaatttagtcaacaataatagtagttagaagaataactgtaatacaata 1460
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 40845 GTTCAATAGATTCATTTAATAATTCATATTTATC-----ATTAATTTATCTTAGAGTTGA 40792
OY 1461 agtgaattatccaacaactctctacgcctttatagcttctcgaagaacttcatc 1520
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 40791 ATTTTGAATATCCAAAAAATGAATTTCTTTTGAAGATATTTCTATTTCTTTGTAAAT 40732
OY 1521 tattcagtaactctttatatacgatactatgtaaatggttccatcattgagcagaatag 1580
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 40731 TGGCAATCAATATGCCATTTTATTTATACATTTCTCGATTTTCTTTTCAAGACTTT 40672
```

```
OY 1581 ttctgttgattatcatcactatcaactgtccatttatactcaagtttcatc 1640
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 40671 CTTCAAGCATATCTTAAAAAGATTAATTAATTCATTTTATTTTCAATTAATTTCTTTTAAAT 40612
OY 1641 tctgataatacaacaataatatacctctgtatctgcccacattttagggcattctccat 1700
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 40611 AGTTTTTGTAAAGATTTGTCATTTTAAATCTTTGGGAAATTTTAAAGCAATTTTATCTTAT 40552
OY 1701 tctgtaacaacttcaacttattgcatattgcaatlaaagaattattcatttgtagaaaa 1760
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 40551 TTTGTAATATTTATTCATCGTTGTTATTAATTCCTCAAGATCAATATATTTTGAATTAATA 40492
```

```
RESULT 12
M60472
ID M60472 standard; DNA: 4590 BP.
AC M60472.
DT 24-AUG-1991 (first entry)
DE Sequence encoding the ring-infected Erythrocyte Surface Antigen
DE (RESA).
KW Malaria vaccine; antigen; epitope; ss.
OS Plasmodium falciparum.
FH Key Location/Qualifiers
FT exon 801..995
FT exon 1199..4225
FT exon /*tag= a
FT exon /*tag= b
FN M08601802-A.
PD 27-MAR-1986.
PR 11-SEP-1985; 006960.
PR 11-SEP-1984; AU-007067.
PR 11-SEP-1984; AU-007066.
PR 10-SEP-1985; AU-047326.
PA (HALL-) HALU INST MED RES.
PI Kemp DJ, Anders R, Coppel RL, Brown G, Saint RB, Cowman AF,
PI WPI: 86-094065/14.
DR P-PSDB: P60569.
PT DNA coding for Plasmodium falciparum antigens - expressing
PT poly:peptide(s) having antigenicity of RESA or FIRA antigens of P
PT falciparum
PS Claim 4; Fig 1; 55pp; English.
CC The inventors claim a novel DNA molecule which comprises a
CC nucleotide sequence corresp. to all or a portion of the base
CC sequence coding RESA (M60472) or FIRA (M60473). RESA and FIRA have
CC antigenicity suitable for providing protective immunity against
CC Plasmodium falciparum malarial infections.
SQ Sequence 4590 BP; 1933 A; 437 C; 673 G; 1547 T;
```

```
Query Match 2.8%; Score 50.6; DB 1; Length 4590;
Best Local Similarity 46.2%; Pred. No. 0.094;
Matches 203; Conservative 0; Mismatches 234; Indels 2; Gaps 1;
```

```
OY 1339 ttagaatcatcgaagcaataataaactgcacatatatactgttcttcttcatcctt 1398
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 304 TAAATATCAAAAAAATAAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 363
OY 1399 ccaataaccaatttagctcaacaataatagtagtaggaataaattactctgatacaa 1458
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 364 TTTTATTTTATTTTATTTTATTTATTTATTTATTTTGTGTAAGAAAAAATAAATAAATAA 423
OY 1459 taagtagataatttcacaaactctctcatcgcctttagcttctcgaagaactatcc 1518
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 424 AAAAAATATATTTATTTATTAATATTAATTAATTAATTTTATTTTATTTATTTATTTATGATAGCG 483
OY 1519 attatcagtaacttcttatacgatactatgtaaatggttccatcattcagcagtaat 1578
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 484 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 541
OY 1579 agttcgttggttatatacatcaactacgaactgtccacttcttattctcaagtttata 1638
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 542 ATTTATATATTTATTTTATTTTATTTTATTTTATTTATTTATTTATTAATTAATTAATTAATTA 601
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: April 16, 2000, 01:53:46 ; Search time 165.68 Seconds
(without alignments)
1285.535 Million cell updates/sec

Title: US-09-323-427-1

Perfect score: 1779

Sequence: 1 ggttaattaccacaagtty.....aaaaaaaaaaaaaaaaaaaa 1779

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 214294 seqs, 59861574 residues

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/5C.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/5D.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/6.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/PCTUS9.COMB.seq: *
7: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64	3.6	665	4 US-08-883-795A-36	Sequence 36, Appl
2	57.8	3.2	6243	4 US-09-056-075-1	Sequence 1, Appl1
3	55.4	3.1	19124	4 US-08-487-826B-13	Sequence 13, Appl
4	54.6	3.1	5852	1 US-07-867-106-2	Sequence 2, Appl1
5	54	3.0	8920	3 US-08-446-855A-1	Sequence 1, Appl1
6	51.6	2.9	665	4 US-08-883-795A-36	Sequence 36, Appl
7	51	2.9	19124	4 US-08-487-826B-13	Sequence 13, Appl
8	49.6	2.8	7218	1 US-08-232-463-14	Sequence 14, Appl
9	48.6	2.7	8920	3 US-08-446-855A-1	Sequence 1, Appl1
10	47.4	2.7	4673	1 US-07-638-431-1	Sequence 1, Appl1
11	47.4	2.7	2504	6 PCT-US92-00018-1	Sequence 1, Appl1
12	47	2.6	2504	1 US-08-484-105-15	Sequence 15, Appl
13	47	2.6	2504	1 US-08-484-106-15	Sequence 15, Appl
14	46.4	2.6	9636	2 US-08-323-170B-1	Sequence 1, Appl1
15	46.2	2.6	5181	2 US-08-257-073-10	Sequence 10, Appl
16	45.6	2.6	1939	1 US-07-715-751B-2	Sequence 2, Appl1
17	44.6	2.5	5852	1 US-07-867-106-2	Sequence 2, Appl1
18	44.4	2.5	1422	2 US-08-319-704-5	Sequence 5, Appl1
19	43.6	2.5	1435	1 US-08-302-449-1	Sequence 1, Appl1
20	43.6	2.5	1435	6 PCT-US94-07430-1	Sequence 1, Appl1
21	43.6	2.5	6243	4 US-09-056-075-1	Sequence 1, Appl1
22	43.4	2.4	4818	5 US-08-817-926-27	Sequence 27, Appl
23	43.2	2.4	198	6 PCT-US95-10668-1	Sequence 1, Appl1
24	43.2	2.4	198	6 PCT-US95-10668-2	Sequence 2, Appl1
25	43.2	2.4	198	6 PCT-US95-10668-3	Sequence 3, Appl1
26	43.2	2.4	198	6 PCT-US95-10668-4	Sequence 4, Appl1
27	43	2.4	2570	4 US-09-056-075-2	Sequence 2, Appl1

28	42.6	2.4	1493	7 5340934-5	Patent No. 5340934
29	42	2.4	731	2 US-08-451-405A-2	Sequence 2, Appl1
30	42	2.4	921	4 US-08-795-475-2	Sequence 2, Appl1
31	42	2.4	1308	4 US-08-795-475-4	Sequence 4, Appl1
32	42	2.4	3607	2 US-08-647-351B-1	Sequence 1, Appl1
33	42	2.4	4766	6 PCT-US93-07261-10	Sequence 10, Appl
34	42	2.4	10136	1 US-08-353-700-2	Sequence 2, Appl1
35	42	2.4	10136	6 PCT-US95-16216-2	Sequence 2, Appl1
36	41.6	2.3	1689	4 US-07-991-867B-41	Sequence 41, Appl
37	41.6	2.3	1689	4 US-08-544-332-41	Sequence 41, Appl
38	41.6	2.3	3459	4 US-08-980-060-3	Sequence 3, Appl1
39	41.6	2.3	8457	1 US-07-991-867B-1	Sequence 1, Appl1
40	41.6	2.3	8457	4 US-08-544-332-1	Sequence 1, Appl1
41	41.2	2.3	1511	1 US-07-991-867B-8	Sequence 8, Appl1
42	41.2	2.3	1511	2 US-08-107-755A-8	Sequence 8, Appl1
43	41.2	2.3	1511	4 US-08-544-332-8	Sequence 8, Appl1
44	40.8	2.3	731	2 US-08-451-405A-2	Sequence 2, Appl1
45	40.4	2.3	660	1 US-07-991-867B-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1
US-08-883-795A-36
; Sequence 36, Application US/08883795A
; Patent No. 5985607
; GENERAL INFORMATION:
; APPLICANT: Delcove, Genevieve
; ATTORNEY/AGENT INFORMATION:
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
; TITLE OF INVENTION: Vectors for Tissue Plasmidogen Activator
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/883,795A
; FILING DATE: 27-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravalle, Micheline
; REGISTRATION NUMBER: 40,261
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 665 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: Rh 32
; US-08-883-795A-36
Query Match 3.6%; Score 64; DB 4; Length 665;
Best Local Similarity 47.8%; Pred. No. 4.3e-06;
Matches 217; Conservative 0; Mismatches 235; Indels 2; Gaps 1;

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 19124 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHEICAL: NO
 ANTI-SENSE: NO
 US-08-487-826B-13

Query Match 3.1%; Score 55.4; DB 4; Length 19124;
 Best Local Similarity 47.1%; Pred. No. 0.001;
 Matches 170; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

QY 1419 acaatatatagtagtaagaataatctgtaatacacaatagatgatatcttcacaa 1478
 DB 1231 ATATATATATATATAGTATAGTATCAAAATATATATATATATATATATATAT 1290
 QY 1479 acttcctatcgtcttataagctcctgaagaagcttattcattcagtaacattat 1538
 DB 1291 TATATTTGTAACATACACAACTAAGAAACATATACATCTGATCTATAGTATAT 1350
 QY 1539 atgcacacattgtaagtgatgcatcattgacgcaatgaaagtgatgctgttatatca 1598
 DB 1351 ATATATATATATCTTTTATATATATATATATATATATATATATATATATATAT 1410
 QY 1599 tcaatcacactgctccttatttcttcaacagttatcattgcatgataatccaaat 1658
 DB 1411 ATATATATATATTTTTCATATATATATATATATATATATATATATATATATAT 1470
 QY 1659 tataccgtgattgcccaattttatggcgcattcctcattcgtgaacaattcact 1718
 DB 1471 TTTAAATACTTCAAAACATTTTGCATAAATAATATATATATATATATATATAT 1530
 QY 1719 attgcattatgcaatttaaaagattcattcttgtaaaaaaataaaaaaataaa 1778
 DB 1531 AATATGAGAGAAACGTAGAACATACCAAAAAAATATAGAACAAAAAGATATATAC 1590
 QY 1779 a 1779
 DB 1591 A 1591

RESULT 4
 US-07-867-106-2
 Sequence 2, Application US/07867106
 Patent No. 5389526

GENERAL INFORMATION:
 APPLICANT: Slade, Martin B
 APPLICANT: Chang, Andy C M
 APPLICANT: Williams, Keith L
 TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
 TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
 STREET: One Liberty Place 46th Floor
 CITY: Philadelphia
 STATE: PA
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/867.106

FILING DATE: 19920625
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: AU PJ 7187
 APPLICATION NUMBER: PCT/AU90/00530
 FILING DATE: 02-NOV-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Feeney, Joanne Longo
 REGISTRATION NUMBER: 35,134
 REFERENCE/DOCKET NUMBER: RICE-0002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-568-3100
 TELEFAX: 215-568-3439
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 5852 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ANTI-SENSE: NO
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 2378..5038
 NAME/KEY: CDS
 LOCATION: 2378..5038
 US-07-867-106-2

Query Match 3.1%; Score 54.6; DB 1; Length 5852;
 Best Local Similarity 46.8%; Pred. No. 0.0011;
 Matches 214; Conservative 0; Mismatches 234; Indels 9; Gaps 1;

QY 1332 aataatgtagatcatcgaagaacaataaataacgcatatataatgcttcttctat 1391
 DB 1583 AAAAAATATAAAAAATACTAGTATATATATATATATATATATATATATATATAT 1642
 QY 1392 catcttcaataactaatctttagctaaacaatatagtagtaagaataattcgt 1451
 DB 1643 TGTCAATGACACTTTTCTTTTGTGCAATGACACTTTTCTTTTAAAAAATAAATA 1702
 QY 1452 aataaataagtagatatttcaataaacttcttcatcgtcttataagcttcgaag 1511
 DB 1703 GTTAAATATCTATTTTGTGACATTCATTTTCCATGTTTATAGATATATATAT 1762
 QY 1512 ctattcatattcgaatccttataatgcatcactatgtaagttcatcattagc 1571
 DB 1763 TAAATTCCTATCGATATATATATATATATATATATATATATATATATATAT 1822
 QY 1572 catgataagttcgttcttattatcatcattcacttgcctatttcttaacag 1631
 DB 1823 AATTTCTACTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 1882
 QY 1632 ttatcatattgataatcac-----aaattacactglatgccaatttct 1682
 DB 1883 TAGATCTCATATATTAATAATCAATTTAAATTAATAAGTATTTTAAATATGCA 1942
 QY 1683 atgggcatcattcctcattcgttaacaactcactatttgcattatgcaattaaa 1742
 DB 1943 ATAAAAAACCTAATGTATGTTTAAACCAACTTTTCTATTTCTTTTCTTTTCT 2002
 QY 1743 tattcatttgtaaaaaaataaaaaaataaaaaaataaaaaaataaaaaaata 1779
 DB 2003 TTTTACTTTGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2039

RESULT 5
 US-08-446-855A-1
 Sequence 1, Application US/0844685A
 Patent No. 5849573
 GENERAL INFORMATION:
 APPLICANT: Stewart, Thomas S
 APPLICANT: Flores, Maria V

```

; APPLICANT: O'Sullivan, William J
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 1100 No. 5849573th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,855A
; FILING DATE: 06-JUL-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mitchard, Leonard C
; REGISTRATION NUMBER: 29,009
; REFERENCE/DOCKET NUMBER: 47-80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic
US-08-446-855A-1

```

```

Query Match          3.0%; Score 54; DB 3; Length 8920;
Best Local Similarity 47.6%; Pred. No. 0.0017;
Matches 195; Conservative 0; Mismatches 210; Indels 5; Gaps 1;

```

```

QY 1370 atatatatggtttctctcatcatcccttctaataactatlttagtacaataata 1429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 245 AAAAATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 304
QY 1430 gtagtggagaataattactgtaatacaagaatgatttcaacaaacttctctat 1489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 305 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 364
QY 1490 cgcctttatagcttcgaaaagcttattcattcagtaacttcttatacgatactat 1549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 365 TTATTAATTAATCTGAAATATTAACCTTATTAATATTTCCAAATTAATATGAAATCAAT 424
QY 1550 tgaataagttcatcatcagcagataggttcgttgtttatcatcatcatcaac 1609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 425 TATTAATATTTTGTGTCACATTAATATAGTT----TTACACTCTTATTAATATAAA 479
QY 1610 ttgtccattttatctcaacagtttaccatttgataatcacaaattataacttgta 1669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 480 CCATCCCATATATATACACATATATATACCCCAATATGTGTCTCCATTAATTTT 539
QY 1670 ttgcccattttatagggcatcttccatactcgttaaacattcacttattgcatat 1729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 540 ATTATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAAT 599
QY 1730 tgcgaataaagaattcatttgtgaaaaaataaataaataaataaataa 1779
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 600 AGTAATTTCTACTAATTTAAAAAATAAATAAATAAATAAATAAATAAAGAAA 649

```

```

RESULT 6
US-08-883-795A-36/C
; Sequence 36, Application US/08883795A

```

```

; Patent No. 5985607
; GENERAL INFORMATION:
; APPLICANT: Delcive, Genevieve
; APPLICANT: Awang, Gregor
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/883,795A
; FILING DATE: 27-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Micheline
; REGISTRATION NUMBER: 40,261
; REFERENCE/DOCKET NUMBER: 7841-062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 665 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: Rh 32
US-08-883-795A-36

```

```

Query Match          2.9%; Score 51.6; DB 4; Length 665;
Best Local Similarity 49.6%; Pred. No. 0.0028;
Matches 222; Conservative 0; Mismatches 214; Indels 12; Gaps 3;

```

```

QY 1328 taataataatgtagaatacatcgaagaataataaactgcca--tatattcgtttct 1385
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 463 TATTAATAATATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 404
QY 1386 tcttcatcatctcttaataactaattttagcacaatatatagtagtaggaataat 1445
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 403 TATTAACATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 344
QY 1446 tactgtaacaacaagtgatatttcatcacaactctcttcatcgcttataagcttct 1505
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 343 TATTAACATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 284
QY 1506 gaaagcttattcattatcagtaactcttataatgatactactatgt--aaatggttcat 1563
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 283 TATTAACATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 224
QY 1564 catlaggcagataggttcgttgttatatcatcatcatcacttgctcattttat 1623
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 223 TATTAATATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 164
QY 1624 tctaacagtttaccatttgtagataatatacaacaatta-----tacttgatggccc 1675
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 163 TATTAATATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 104
QY 1676 aattttatgggcatcatttccctatctcgtgtaacaattcacttattgcatattgcaat 1735

```


Db 1096 YY 1155
QY 1425 atagatagtagaataactctgaataagatatacttcacaaactctc 1484
Db 1156 YY 1215
QY 1485 tcatagccttataagctcgaagaactatcatcattcaglaacttcttatgcac 1544
Db 1216 YY 1275
QY 1545 acatctgtaactgtcatcagcagcagatgagtgctgttatcatcatca 1604
Db 1276 YY 1335
QY 1605 tcaactgtcctatcttatactcaagcttatcatcttgatgaatacacaactac 1664
Db 1336 YY 1395
QY 1665 ttgatgtcccaatttattatgagcatcattctcattctgtaaacattcactattgc 1724
Db 1396 YYGTACCAAACTCTCTATCTCT 1455
QY 1725 attattgc 1732
Db 1456 TTAATACTAC 1463

RESULT 9

US-08-446-855A-1/c
; Sequence 1, Application US/08446855A
; Patent No. 5849573
; GENERAL INFORMATION:
; APPLICANT: Stewart, Thomas S
; APPLICANT: Flores, Maria V
; APPLICANT: O'Sullivan, William J
; TITLE OF INVENTION: Nucleotide sequence encoding carmayol
; TITLE OF INVENTION: phosphate synthetase II
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 1100 No. 5849573th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,855A
; FILING DATE: 06-Jul-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mitchard, Leonard C
; REGISTRATION NUMBER: 29, 009
; REFERENCE/DOCKET NUMBER: 47-80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic
; US-08-446-855A-1

Query Match 2.7%; Score 48.6; DB 3; Length 8920;
Best Local Similarity 49.4%; Pred. No. 0.028;

Matches 126; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
QY 1525 cagtaactcttataatgcactatgtgaatgcttcacatcagtagccatgaatgttc 1584
Db 797 CATATATTTATATATATATATATATATATATATATATCTATATATATTTTCCCATTTTTC 738
QY 1585 gtgtgtatcatcatcatatcaactgtgcctattcttatacgaagttatcatcttg 1644
Db 737 TTTTATTTTATATACATTTATATATTTATGTATTAATATTTTAAATTTTACATATACAAG 678
QY 1645 ataatatcaccaatataactctgtatgtcccaattttataggcactatctctctc 1704
Db 677 TTCATTTTTCATATATGAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 618
QY 1705 taacaattcactatttgcattatgtcaatcaataaagaattcatttgtagaataaaa 1764
Db 617 TAAATTAGTAGAATTAAGTATTTTATATACTAAGAAAAAATAAATAATGAATATAATTA 558
QY 1765 aaaaaaaaaaaaaa 1779
Db 557 ATAAATAAATATATATA 543

RESULT 10

US-07-638-431-1
; Sequence 1, Application US/07638431
; Patent No. 5198535
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Chareonvilt, Yupin
; APPLICANT: Hedstrom, Richard
; APPLICANT: Khushmith, Srisin
; APPLICANT: Rogers IV, William O.
; TITLE OF INVENTION: Protective malaria sporozoite surface protein
; TITLE OF INVENTION: immunogen and gene
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: A. David Spevack
; STREET: NMDC Building 1 T-12 National Naval
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20814-5044
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/638,431
; FILING DATE: 19910110
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, Avrom D.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-4033
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4673 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: N
; ANTI-SENSE: N
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium yoelii
; STRAIN: 17X(NL)
; DEVELOPMENTAL STAGE: erythrocytic stage
; TISSUE TYPE: Blood
; CELL TYPE: erythrocytic stage

IMMEDIATE SOURCE:
LIBRARY: Py-lambda-gl1-2-7 kb genomic expression
CLONE: Py10.1111
FEATURE:
NAME/KEY: CDS
LOCATION: 718..3195
OTHER INFORMATION:
US-07-638-431-1

Query Match 2.7%; Score 47.4; DB 1; Length 4673;
Best Local Similarity 48.1%; Pred. No. 0.043;
Matches 165; Conservative 0; Mismatches 176; Indels 2; Gaps 1;

QY 1439 aaataatcagtaataacataagatgatatcttcacaaactctctcatgcttcat 1498
DB 3279 AAATATTATCTATGATTAATATATATATATACCTTTAAATAATATCATAAATCGCT 3338
QY 1439 agcttcgaaagctatcatcattcagtaactcttatacgatactatgtaagt 1558
DB 3339 TGCTTTAAATGTTTGTTCTTTACACTTATTCCTTTTCCCTGTTTGTCCTTT 3398
QY 1559 ttcatcattagggcgaatagttcgttctgttattatcatcattacaacttgctcat 1618
DB 3339 TTTTGTATGATTAAGTATTTAAATTAACAGTTTGATTAATGTCATCTTTTATGT 3458
QY 1619 --ttatctcagatgattcattgttgataataacacaaattacacttgatgccca 1676
DB 3459 TATTCATTCATTAATTAATATCCATTTATTTTAAACGATTTTAAAC 3518
QY 1677 attttaaggacatattccattctgtaacaattcactattctgcatattgcaatt 1736
DB 3519 TATTTTATTAACATTAATGTCGTGTAATATATATATATTTATTTATACCTCAATATT 3578
QY 1737 aaaaagtattcattctgtaaaaaaataaaaaaataaaaaa 1779
DB 3579 AATGCTTACATTAATTTCTTAATATATAAAAAA 3621

RESULT 11

Sequence 11, Application PC/TUS9200018
PCT-US92-00018-1
GENERAL INFORMATION:
APPLICANT: Hoffman, Stephen L.
APPLICANT: Charoenyit, Yupin
APPLICANT: Hedstrom, Richard
APPLICANT: Khushf, Srisin
APPLICANT: Rogers IV, William O.
TITLE OF INVENTION: Protective malaria sporozoite surface protein
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: A. David Spevack
STREET: NMDC Building 1 T-12 National Naval
CITY: Bethesda
STATE: MD
COUNTRY: USA
ZIP: 20814-5044
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00018
FILING DATE: 19920103
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, Avram D.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-6759
TELEFAX: (301) 295-4033

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4673 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
ORIGINAL SOURCE:
ORGANISM: Plasmodium yoelii
STRAIN: 17X(NL)
DEVELOPMENTAL STAGE: erythrocytic stage
TISSUE TYPE: Blood
CELL TYPE: erythrocytic stage
IMMEDIATE SOURCE:
LIBRARY: Py-lambda-gl1-2-7 kb genomic expression
CLONE: Py10.1111
FEATURE:
NAME/KEY: CDS
LOCATION: 718..3195
OTHER INFORMATION:
PCT-US92-00018-1

Query Match 2.7%; Score 47.4; DB 6; Length 4673;
Best Local Similarity 48.1%; Pred. No. 0.043;
Matches 165; Conservative 0; Mismatches 176; Indels 2; Gaps 1;

QY 1439 aaataatcagtaataacataagatgatatcttcacaaactctctcatgcttcat 1498
DB 3279 AAATATTATCTATGATTAATATATATATATACCTTTAAATAATATCATAAATCGCT 3338
QY 1499 agcttcgaaagctatcatcattcagtaactcttatacgatactatgtaagt 1558
DB 3339 TGCTTTAAATGTTTGTTCTTTACACTTATTCCTTTTCCGTTTGTCCTTT 3398
QY 1559 ttcatcattagggcgaatagttcgttctgttattatcatcattacaacttgctcat 1618
DB 3399 TTTTGTATGATTAAGTATTTTAAATTAACAGTTTGATTAATGTCATCTTTTATGT 3458
QY 1619 --ttatctcagatgattcattgttgataataacacaaattacacttgatgccca 1676
DB 3459 TATTCATTCATTAATTAATCCATTTATTTTAAACGATTTTAAAC 3518
QY 1677 attttaaggacatattccattctgtaacaattcactattctgcatattgcaatt 1736
DB 3519 TATTTTATTAACATTAATGTCGTGTAATATATATATTTATTTATACCTCAATATT 3578
QY 1737 aaaaagtattcattctgtaaaaaaataaaaaaataaaaaa 1779
DB 3579 AATGCTTACATTAATTTCTTAATATATAAAAAA 3621

RESULT 12

Sequence 12, Application US-08-484-105-15/c
US-08-484-105-15/c
Patent No. 5589341
GENERAL INFORMATION:
APPLICANT: STILLMAN, Bruce
APPLICANT: BELL, Stephen P
APPLICANT: KOBAYASHI, Ryuji
APPLICANT: RINE, Jasper
APPLICANT: FOSS, Margit
APPLICANT: MCNALLY, Francis J
APPLICANT: LAURENSEN, Patricia
APPLICANT: HERSKOWITZ, Ira
APPLICANT: LI, Joachim J
APPLICANT: GAVIN, Kimberly
TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT


```
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/323.1708
  FILING DATE: 13-OCT-1994
  CLASSIFICATION: 424
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/010,409
    FILING DATE: 29-JAN-1993
    ATTORNEY/AGENT INFORMATION:
      NAME: Quine, Jonathan A.
      REGISTRATION NUMBER: P-41,261
      REFERENCE/DOCKET NUMBER: 015280-11310005
      TELECOMMUNICATION INFORMATION:
        TELEPHONE: (415) 576-0200
        TELEFAX: (415) 576-0300
      INFORMATION FOR SEQ ID NO: 1:
        SEQUENCE CHARACTERISTICS:
          LENGTH: 9636 base pairs
          TYPE: nucleic acid
          STRANDEDNESS: single
          TOPOLOGY: linear
        MOLECULE TYPE: DNA (genomic)
        FEATURE:
          NAME/KEY: CDS
          LOCATION: 149..9556
US-08-323-1708-1

Query Match      2.6%; Score 46.4; DB 2; Length 9636;
Best Local Similarity 46.4%; Pred. No. 0.09;
Matches 257; Conservative 0; Mismatches 291; Indels 6; Gaps 3;

QY 1214 aatgaatcgtgcatgaccattgcttcacatgattatgaggttaagcattgcatg 1273
DB 7021 AATAACATCTTTAGGTTTAATTAATTTTATCTTTATGCAATTAATTAATTAATG 6962
QY 1274 attgctgcgcgtcatattaccattcgtttaaatctgcaccaatcagaagggcataaaa 1333
DB 6961 TTCACATCTCTCATATCATCTTTTAATTAATTCATAAACAATTAACATGACCTTG 6902
QY 1334 t--aatgttagaatcatcgaagaacataaataaacgcctatatattcgcttctctat 1391
DB 6901 TGGAAATATAGCAAAACCTTAGACATTTCATCCAAAATAATCTCTTTTAAATTTT 6842
QY 1399 catcctcttaactaactaatttagcttaacaatatatagtagtaagaaataactatgt 1451
DB 6841 AACAGTGCACCTTTTCTTTTACTTGACTATTTATTCATTATGAAGTGCAGACCTCT 6782
QY 1452 aatacaataagtagataatttcatacaaacctctctcctatcgctttatagcttcgaaaag 1511
DB 6781 TGTGAATCA--CATTAATTTAATGAATTTCTGATACATTTTCTCTTAACAAAATAATAT 6724
QY 1512 ctattcctattcagtaactcttta--tagcatactatgtaaatggtttcatcatag 1569
DB 6723 ATTAATTCCTCTTTGTTTAATTAATTTGATTAATCTATTGTTTCAATTAATTAATTTTAA 6664
QY 1570 gccatgaatagttcgttgttatatcatcatcaactgctccatttattcctaaac 1629
DB 6663 TCTGATCTCTGTTTATGAGATTAATAAAACATAATTCCTTCCTTTTTCATTTT 6604
QY 1630 agttatcatctgtgataataatcacanaattacacctgttatgcccnaattttatgggca 1689
DB 6603 TTATCAAAAGTAAATAATTCGATTTCAATTAACCCCTTCAAGCTAAACTAATATAGTACT 6544
QY 1690 taatttcctattctgtaacaattcaactatttgcatattgccaatlaaaagatttca 1749
DB 6543 TCATTTTCTTTCTTTTCTTAATCACTTAATATCATTAATGTTATTCATCAAAAACATAATGAT 6484
QY 1750 ttctgtaaaaaaaa 1763
DB 6483 GGTTTTAATGAAAA 6470
```

```
RESULT 15
US-08-257-073-10
Sequence 10, Application US/08257073
Patent No. 5766597
GENERAL INFORMATION:
  APPLICANT: Paoletti, Enzo
  APPLICANT: de Taisne, Charles
  APPLICANT: Tine, John A.
  TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
  NUMBER OF SEQUENCES: 143
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Curtis, Morris & Safford, P.C.
    STREET: 530 Fifth Avenue, 25th Floor
    CITY: New York
    STATE: New York
    COUNTRY: UNITED STATES OF AMERICA
    ZIP: 10036
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/257,073
    FILING DATE: 09-JUN-1994
    CLASSIFICATION: 424
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 08/075,783
      FILING DATE: 11-JUN-1993
      PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 07/852,305
        FILING DATE: 18-MAR-1992
        APPLICATION NUMBER: US 07/672,183
        FILING DATE: 20-MAR-1991
        ATTORNEY/AGENT INFORMATION:
          NAME: Frommer, William S.
          REGISTRATION NUMBER: 25,506
          REFERENCE/DOCKET NUMBER: 454310-2570
          TELECOMMUNICATION INFORMATION:
            TELEPHONE: (212) 840-3333
            TELEFAX: (212) 840-0712
            TELEX: 425066 CURTMS
          INFORMATION FOR SEQ ID NO: 10:
            SEQUENCE CHARACTERISTICS:
              LENGTH: 5181 base pairs
              TYPE: nucleic acid
              STRANDEDNESS: single
              TOPOLOGY: linear
US-08-257-073-10

Query Match      2.6%; Score 46.2; DB 2; Length 5181;
Best Local Similarity 52.3%; Pred. No. 0.083;
Matches 102; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 1576 aatgattcgttgttatattcatcatcaactgctccatttattcctaagttta 1635
DB 3046 AATGATACCTATTTTATATGATATCAACAATTCGTAAATCTTAATTCAAAAGTAATT 3105
QY 1636 taatttgataataatacaaatatcaccttgatgcccnaattttatgggcatcaatt 1695
DB 3106 ACAGGTTTACCGAACAACAAAATAATGCATTAAATGATGAATTAATAAATTAAGAAT 3165
QY 1696 cctattctgtaacaattcaactatttgcatattgccaatlaaaagatttcaattgtg 1755
DB 3166 ACTTTAAGTTATCATTTGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3225
QY 1756 aaaaaaaaaaaaaa 1770
DB 3226 AATAAGAAAAAGAA 3240
```

Thu Apr 17 11:11:09 2003

us-09-323-427-1.rni

Page 10

Search completed: April 16, 2000, 04:35:02
Job time: 9676 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 16, 2000, 01:25:56 ; Search time 3225.5 Seconds
(without alignments)
2082.439 Million cell updates/sec

Title: US-09-323-427-1
Perfect score: 1779
Sequence: 1 ggttaattaccacgaatttg.....aaaaaaaaaaaaaaaaaaaaa 1779

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*

45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	391.8	22.0	665	42	A111196 SMOV3MCA1
2	373.8	21.0	628	37	AA701731 SMOV3MCA1

3	369.4	20.8	639	44	A1322117	A1322117	SMOV3MCA1
4	365.8	20.6	672	36	AA618895	SMOV3MCA1	
5	355.8	20.0	610	37	AA668071	SMOV3MCA1	
6	337.2	19.0	776	35	AA585626	SM3D9CA34	
7	314	17.7	537	37	AA668051	SMOV3MCA1	
8	308.4	17.3	552	44	A1322068	SMOV3MCA1	
9	301.2	16.9	615	44	A1317885	SMOV3MCA1	
10	292.8	16.5	531	39	AA901444	SMOV3MCA1	
11	278.8	15.7	363	36	AA625020	SMOV3MCA1	
12	269.2	15.1	466	36	AA625024	SMOV3MCA1	
13	269.2	15.1	354	39	AA841200	MB3D6AA4G	
14	264.6	14.9	459	36	AA618952	SMOV3MCA1	
15	260.4	14.6	395	36	AA625010	SMOV3MCA1	
16	257	14.4	369	36	AA625022	SMOV3MCA1	
17	248.4	14.0	528	44	A1322078	SMOV3MCA1	
18	241.2	13.6	375	36	C71078	SMOV3MCA1	
19	237.2	13.3	379	62	AV203892	AV203892	
20	235.8	13.3	360	36	C67783	SMOV3MCA1	
21	232.4	13.1	377	36	C70345	SMOV3MCA1	
22	223.8	12.6	493	74	AA208357	SMOV3MCA1	
23	221.4	12.4	504	36	AA624955	SMOV3MCA1	
24	215	12.1	360	62	AV201949	AV201949	
25	210.8	11.8	360	62	AV196275	AV196275	
26	210.2	11.8	377	62	AV186785	AV186785	
27	207.2	11.6	360	62	AV202546	AV202546	
28	205.4	11.6	322	36	AA625021	SMOV3MCA1	
29	205.6	11.5	360	62	AV201206	AV201206	
30	204.8	11.5	360	62	AV191120	AV191120	
31	193.6	10.9	417	36	AA625044	SMOV3MCA1	
32	173.2	9.7	443	36	AA625040	SMOV3MCA1	
33	163.2	9.2	281	46	A138781	SMOV3MCA1	
34	162.2	9.1	318	20	T01970	WEST02691 E	
35	161	9.1	380	44	A1322108	SMOV3MCA1	
36	151.4	8.5	635	44	A1317939	SMOV3MCA1	
37	149.2	8.4	522	44	A1322073	SMOV3MCA1	
38	143.4	8.1	467	31	AA294218	SMOV3MCA1	
39	140.2	7.9	317	45	A1352993	SMOV3MCA1	
40	119	6.7	273	49	A165774	SMOV3MCA1	
41	109.2	6.1	473	36	AA625005	SMOV3MCA1	
42	102.6	5.8	499	44	A1313768	SMOV3MCA1	
43	100.6	5.7	470	44	A1313804	SMOV3MCA1	
44	89.4	5.0	409	36	AA625029	SMOV3MCA1	
45	86.2	4.8	407	36	AA625007	SMOV3MCA1	

ALIGNMENTS

RESULT 1
A111196 665 bp mRNA EST 31-AUG-1998
LOCUS SMOV3MCA1232SR Onchocerca volvulus molting L3 larva cDNA
DEFINITION (SL96MLW-OvML3) Onchocerca volvulus cDNA clone SMOV3MCA1232 5',
mRNA sequence.

ACCESSION A111196
VERSION A111196.1 GI:3510080
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spirurida;
Filarioidea; Onchocercidae; Onchocerca.

REFERENCE 1 (bases 1 to 665)
AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL Unpublished (1997)

COMMENT On Sep 12, 1996 this sequence version replaced gi:1407460.

Contact: Steven A. Williams
Molecular Parasitology

Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA

Tel: 4135853826
Fax: 4135853786

FEATURES	source	Email: genome@smith.edu Seq primer: pbluescript SK. Location/Qualifiers
1. 665		/organism="Onchocerca volvulus" /strain="Kumba, Cameroon" /db_xref="taxon:6282" /clone="SMOV3MCA1232" /clone_lib="Onchocerca volvulus molting L3 larva cDNA (SL96MLW-OvML3)" /dev_stage="molting L3" /lab_host="X11-Blue MRF" /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (forest strain). The L3 were cultured in 20% FCS in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volvulus molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (mL3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda Uni-Zap XR vector and has 1 x 10E6 independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustig@nbc.org)."
BASE COUNT	186 a 129 c 136 g 213 t	1 others
ORIGIN		
Query Match	22.0%	Score 391.8; DB 42; Length 665;
Best Local Similarity	79.9%	Pred. No. 1,1e-68;
Matches 473; Conservative 0; Mismatches 118; Indels 1; Gaps 1;		
QY	160	tcttgaatgatgatcgtctctatcgtcttctgcttacctatgatcgtctctatc 219
DB	61	TACTGATATTATCATCTTCGCTTATTAAGCGTATGCTATTATCATCGCTAATACG 120
QY	220	gatttcggttgcacatggtgctggaagtgagcgaagaattgattgagccacttacc 279
DB	121	TATTCGGTTGTAACGCGTGTAGAACGTAACCAAAATTAATGTGTCACATCAAT 180
QY	280	aacaatcaatttaatacaacgtaatgcatacgaagaacatgtttatgtaaaagttcta 339
DB	181	TACTGCAACTTTAATACCTGTAATCATTTTGAAGACATGTATAGTAAAGCTTATA 240
QY	340	tgaatcaagaagtggtccgtaatgaaagtgagcgtcaagttgcccgaatttacc 399
DB	241	CGATCAGGACGAGTGGCGAAATGATGAGGTGACGTACGCGGAATGAATGAACTTCC 300
QY	400	atttgattatcgaatgctgctgctgacagatctcgaatcgaatgctgctgctgctgac 459
DB	301	GTTTGATTTCTGTAATGTAGACACGTACAGCTTCGTTAAATCAACGCGTATTTTGTCCAC 360
QY	460	aacaactgttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgc 519
DB	361	ATCAGTTGTTGTCATCTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 420
QY	520	agatacattgcttatacgaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaa 579
DB	421	AATACATGCTTTTACATGAGGTGATGAGGTGATGAGGTGATGAGGTGATGAGGTGATG 480
QY	580	aatacaactgtcttcaaacatcaatgtcccgatgacagatgacagatgacagatgac 639
DB	481	AATGACAACTGATTTGCAACAAATTTGTCGAGTCCGATGCGATGAGATGAGATGAG 540
QY	640	gagatggtgac 699
DB	541	TGATGTGACCACTCCGACAACTGCTTCAATTTGCTACCAATTTGCTACCACTGCT-CCA 599

```

Oy      700      taatgagcagcgatcttgaaacggtgatctcttcgcgggtgtccat 751
          |||||
Db      600      TAAATGACATGTGATTCTGAAACAGTTGATACATTCTGGCTGGTGACAT 651
          |||||

RESULT  2
LOCUS   AA701731
DEFINITION
SMOY3MCA1144SK Onchocerca volvulus molting L3 larva cDNA
(S1569M1M-Ovm3K) Onchocerca volvulus CDNA clone SMOY3MCA1144 5',
mRNA sequence.
AA701731
ACCESSION
VERSION AA701731.1 GI:2704931
KEYWORDS
SOURCE  EST.
        Onchocerca volvulus.
        Onchocerca volvulus.
        Euryptora; Metazoa; Nematoda; Secernentea; Spirurida;
        Filarioidea; Onchocercidae; Onchocerca.
REFERENCE
AUTHORS Williams,S.A., Litzote-Wanlewski,M., laney,S. and Lustigman,S.
TITLE    Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL  Unpublished (1997)
COMMENT  On Sep 19, 1997 this sequence version replaced gi:1517220.
        Contact: Steven A. Williams
        Molecular Parasitology
        Smith College Department of Biological Sciences
        Department of Biological Sciences, Clark Science Center, Smith
        College, Northampton, MA, 01063, USA
        Tel: 4135853826
        Fax: 4135853786
        Email: genome@smith.edu
        Seq primer: Bluescript SK.
        Location/Qualifiers
            1..628

```

BASE COUNT	181 a	121 c	125 g	201 t
ORIGIN				

	Query Match	Similarity	21.0%	Score 373.8	DB 37	Length 628
	Best Local	Similarity	79.9%	Pred. No. 4,3e-65		
	Matches	453	Conservative	0	Mismatches 112	Indels 2
						Gaps 1
Qy	160	ttctgaatgaatgcgtgccttattgccttcgctaactaatgcattgcattcattcc	219			
Db	56	TACTGATATTATCATCGCTTCGGGTATATAGCGATATTCGATTATTCACGCTCTAAATGC	115			
Qy	220	gattcgcgttgacaatggtgtcgaaggtgagccagaattgaatgtgacaacattcaat	279			

Db	116	TATCCGGTTGATATACGGTGTAGAAAGGTAAACCAAAATGGATGTGGTCCCAACATCART	175
Oy	280	aacaatcaatttaatacacagtaatgcattcgaaagacatglttaltgtaaaagtcctta	339
Db	176	TACTGTCAACTTATATATCTGTATCCATTTGGAAGGACATGTATACGTAAAGGCTTATA	235
Oy	340	tgtatcaagaaggttgcgctaagtgaaagtggaaactgaagtgcggaattcaacttcc	399
Db	236	CGATCAGGCGAGATGCCGAATGTGTGAAGTGGACGTGAGTGGCCGGAATTGAACTTCC	295
Oy	400	atttgattatcagcaatgtctgcgtacaaagatctcgaatccaaacgtygatttttgaac	459
Db	296	GTTTGATTCCTGTAATGATGAGACAGTACACGTTGTTAAATCCACGTTATTTTGTCCAC	355
Oy	460	aacaacgltgtcaatttcglttcatcatatttgttaccaaagttgacgycgatatcg	519
Db	356	ATTCAGTTGTGTCTATCTCATTTCCATTCATTTGTTGGTGCAAAAGTTGATGAGACATATCG	415
Oy	520	agtcacaaatgctttacaatggaaagctgaataaacagttatgtaacagattgaagtaatcga	579
Db	416	AATCAAAATGCTTTTACATGTGAAGAGCTGATTAAGACAGTTAGGCGTCACTGGAAGTTTCCGA	475
Oy	580	aatcaacaatgcgtcttccaactcaaatatgtccgaatgcagatagccgltatgaaattt	639
Db	476	AATGACAATCTCATTTTCCACACACAAATTTGACCGATGCCGTATGCCGATATGAATTTCT	535
Oy	640	ggaatggttgaccacacgcgtcaacacagttcaatttgtcatatggtgcacgc--agttat	697
Db	536	TGATGTGTGGACATTCGGAGACAACCTGTTCATTTGGTACATTTGGTCAACCCAGTGTAC	595
Oy	698	cataaatgagacatgcgatcttgaaac	74
Db	596	CATAAATGACATGTGATTTTGAATCC	622

FEATURES	source
RESULT 3	
AI322117	
LOCUS	
DEFINITION	AI322117 639 bp mRNA EST 22-DEC-1998
ACCESSION	SMOVC3MCA12G058K Onchocerca volvulus molting L3 larva cDNA
VERSION	(SI196MCA1-Ovml.3) Onchocerca volvulus cDNA clone SMOVC3MCA12G058 5',
KEYWORDS	mRNA sequence.
SOURCE	AI322117
ORGANISM	AI322117.1 GI:4056268
REFERENCE	EST.
AUTHORS	Onchocerca volvulus.
TITLE	Onchocerca volvulus
JOURNAL	Eukaryotes; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;
COMMENT	1 (bases 1 to 639)
	Williams, S.V., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
	Genes expressed in molting L3 larvae of Onchocerca volvulus
	Unpublished (1997)
	On Jan 14, 1998 this sequence version replaced gi:1797212.
	Contact: Steven A. Williams
	Molecular Parasitology
	Smith College Department of Biological Sciences
	Department of Biological Sciences, Clark Science Center, Smith
	College, Northampton, MA, 01063, USA
	Tel: 4135853826
	Fax: 4135853786
	Email: genome@smith.edu
	Seq primer: pb1unescript SK.
	Location/Qualifiers
	1..639

```

FEATURES
source
location/Qualifiers
1..639
/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMO3WCAM12608"
/clone_id="Onchocerca volvulus molting L3 larva cDNA
(SL96MILW-OvmL3)"
/dev_stage="molting L3"
/lab_host="x11-blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site 1: Eco RI; Site 2

```


Qy	400	attgattcatcgaatgctgttcggtacaaagatccctcgaatccaaagctgatttttggtaac	459
Db	343	gTTTgATTCCGTATgATgACgCATCAACgTTCCTTAATgCAAGTGGATTTTGTGCAC	402
Qy	460	aacaacgtgtcattctgcgtttacatatttggtaaccaagttgatcgttatcg	519
Db	403	ATCAGTGTGTTCNNNCATTCATCCATTTGTGTGCACAAATTTGATGAGGATATCG	462
Qy	520	agtaacaatgctttacaatgagaagctgataaacaagttagttcacagaattgaagatctga	579
Db	463	AATACAAATGCTTTTACATGAGAGGTGATTAACACAGTTAAGCCCTCAACTTGAATTTCCGA	522
Qy	580	aatcaacaactgcttttcaaacctcaaatctgctccgattgcgaatgacgttatgcaatttc	639
Db	523	AATGACAACATCTCATTTTGGCNAACAAATGTATGCAGTGCCTGATGCCGATATGAGATTCT	582
Qy	640	ggatggtggtgaaccaaccggttcacaacgaattcaatttgcgtatcatatggttcagccagttatca	699
Db	583	TGATGGTGGACATCCCGGACACACTGTCTCAATTTGCTANCAATGTGCTNACACAGTGTCCA	642
Qy	700	taaatgacatcggatcttctgaaccgttta	729
Db	643	TANATNGANATNTGATTCTCGAAAGTTGA	672

RESULT	5
AA668071	
LOCUS	
DEFINITION	610 bp mRNA EST 20-NOV-1997
ACCESSION	SMOVC3MCA02C01sk Onchocerca volvulus molting L3 larva cDNA
VERSION	(S1966MLU-Ovmn3) Onchocerca volvulus cDNA clone SMOVC3MCA02C01 5'
	mRNA sequence.
AA668071	
AA668071.1	GI:2629570

SOURCE	ORGANISM
Onchocerca volvulus.	Onchocerca volvulus
Onchocerca volvulus	Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spirurida
Onchocerca	Phylum: Nematoda; Class: Secernentea; Order: Spirurida; Family: Filariidae; Genus: Onchocercidae; Species: Onchocerca.
1 (bases 1 to 610)	
Williams,S.A., Licotse-Waniewski,M., Laney,S. and Lustigman,S.	
Genes expressed in molting L3 larvae of Onchocerca volvulus	
Unpublished (1997)	
On Oct 30, 1997 this sequence version replaced gi:2160807.	

```

Email: genome@smith.edu
Seq primer: pbluescript SK.
Location/Qualifiers
1..610
FEATURES
source
```

```

/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMOY3MCAMOZC01"
/clone_1ib="Onchocerca volvulus molting L3 larva cDNA
(SL16KLM-Ovml3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site 1: Eco RI; Site 2:
Xho I; Filarial nematode parasite of humans, third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain).The L3 were cultured in 20% FCS
in RPMI NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H

```

and DNA pol I. The library was constructed in the lambda uni-zap XR vector and has 1 x 10E6 independent recombinants and the average insert size is 1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustigman@bc.ORG).

Query Match	20.0%	Score	355.8	DB	37	Length	610
Best Similarity	74.0%	Pred. No.	1.7e-61				
Match 450; Conservative		0; Mismatches	158;			Indels	0;
						Gaps	0;

QY 326 gtgaaggctttatcatcaagaaggttcgcgtaatgatgaaggtgcagctcaagttgcc 385
||||| ||||| ||| ||| ||| ||||| ||| |||

386 ggaattcactccatttgattcaatgcaatgttgcggtacacagatcttgaatccagt 445

Db 62 GGATTGAAGTTCAGANGGATTTCATGTAAATGTTGAACGATCACGGTCCTTAAAATCCTGCA 121

Db 122 GGTCCTTTTGTATCAACCATTAATTTGTTATTTTCATTTTCATCCGAATTTCAATAATA 181

506 gatcgtgacatacgcgtacacatgcttccacatggaagctgataaacagttagtgcacag 565

566 atgaagatattcgaatcacaaatgctttcaaaatcaaaattgtccagatgccaatgac 625

Db 242 CTGGAAGTATCCGAATGACACACGATTCACAGACACAAGTGGTACCAATGCGCTGTGT 301

302 CGATACGAGATTCTGGAAGGTGGACCACTGGTACACCAATTGGATATGCAATGATCGA 361

686 cagccagttatcataatgacatgcgattcgaacccgtgtatattctctgcgcgrr 745

746 gtcacattccctccttctgcatgatgaacagatgactgtagaaattctgaatgactaat 805

Db 422 GTGCATTCTGTACTGTGATGACGGAAAGTGATGATTACGGTAGAGATTCTGTAATGACAA 481

Db 482 GGTTCGTTGGCAAGATTACTACACAATTTGGATATCATCAATCTATGGT 541
000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
000 ggaagcgcgcccaaaacaaatccgccaaccacaaatccggactccccaacgcgaattcaagtc 000

866 ggcccaagaagctcacgtatacacaatatgcygcgacgcgacacgcctttctatacaatgccag 925

0v 926 atcagat 933
 2b 342 GGCAGGAGACCAATCTTGTGGGACCCAGTCAGCAGCATCTCGAATTTATATTCACCAACCAG TTA

RESULT	6
AA585626	
LOCUS	776 bp mRNA EST 30-DEC-1997
DEFINITION	SW2D9C349SK Brugia malayi L3 molting-day 9 larva cDNA (SWM97MLW-Bml3d9) Brugia malayi CDNA clone SW3D9CA349_5', mRNA sequence.
ACCESSION	AA585626
VERSION	AA585626.1 GI:2393038
KEYWORDS	EST.
SOURCE	Brugia malayi.
ORGANISM	Brugia malayi. Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spirurida; Filarioidea; Onchocercidae; Brugia.

Query Match	17.3%	Score 308.4	DB 44	Length 552
Best Local Similarity	78.5%	Pred. No. 4.9e-52		
Matches 394	Conservative 0	Mismatches 106	Indels 2	Gaps 2
OY	160	ttctgaatgatgatgtcgtcttatctgtcttctgtcctacactatgtcatgtcttattc	219	
Db	49	TACTGATATTATACATGCTTCCTGGTTAATAGCGTATTCGTATTATACACTGCTGTAATAGC	108	
OY	220	gattccggtttgcaaatggtgtcgaaggtgagccagaatgtgaatgtggaccaactcaat	279	
Db	109	TATTCGGGTGTAACGGGTGTGAAGGTGAACCAAAATTGAATGTGGTCCAACTCAT	168	
OY	280	aacaatcaatttaacacgtaatgatcgtcgaagacatglttaatgtgaaagtcctta	339	
Db	169	TACTGCTCAACTTTATATACCTGTAATCCATTGTAAGAGCATGTATACGTGAAGAGCCTTATA	228	
OY	340	tgatcaagaaggttgcgtaataatgaaggtgagacgtcaagttgccggaattcacttc	399	
Db	229	CGATAGCGAGGATGCCAAATGATGAAGGTGAGAGTGAAGCGGGAATTAAGCACTTCC	288	
OY	400	attgatcatcgaatgtgttcgctgacacgactctctgtaatccacgttggtattttgtaac	459	
Db	289	GTTTATTTCTCTTAATGTAGACAGCTACACGTTCTTTAATTCACGTTGATTTTTGTCTAC	348	
OY	460	aacaactgttcattcgttttcatcattatgtttaccaaagttgacgtgcatatcg	519	
Db	349	ATCACTGTTGTCATCTCATTCATTCACATTTGTTGTGACAAAGTTGATCGACATATCG	408	
OY	520	agtcacaatgctttacatagtaagctgtataaaacagttggtgcacagattgaagat-ctg	578	
Db	409	AATFACAATGCTTTTACATGAGAGCTGATTAAGACAGTATGAGCGCTCACTGAAGTTCCCG	468	
OY	579	aaatacaacatgcttttcaactcaaatgttccgcatggcagta-tgcgltataaatt	637	
Db	469	AAATACAACTGTCATTGTGCACCAAAATTGTACGATGCTGTATTGGCGGATATGAGATT	528	
OY	638	ttgagatgtgagacaacggtc	659	
Db	529	CTTGATGTGTGAGACATCCCGGC	550	
RESULT	9			
LOCUS	AI317885	615 bp	mRNA	EST
DEFINITION	SMOV3MCA06H12SK Onchocerca volvulus molting L3 larva cDNA (S196H1M-Ovm13)			17-DEC-1998
ACCESSION	AI317885			
VERSION	AI317885.1	GI:4033152		
KEYWORDS	EST.			
SOURCE	Onchocerca volvulus.			
ORGANISM	Onchocerca volvulus; Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;			

KEYWORDS	EST.
SOURCE	Onchocerca volvulus.
ORGANISM	Onchocerca volvulus.
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spirurida;
AUTHORS	Filarioida; Onchocercidae; Onchocerca.
TITLE	1 (bases 1 to 466)
JOURNAL	Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
COMMENT	Genes expressed in molting L3 larvae of Onchocerca volvulus Unpublished (1997) On May 9, 1995 this sequence version replaced gi:802407. Contact: Steven A. Williams Molecular Parasitology Smith College Department of Biological Sciences Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA Tel: 4135853826 Fax: 4135853786 Email: genomsmith.edu Seq primer: pluscript SK.
FEATURES	Seq primer: pluscript SK.
source	Location/Qualifiers 1. 466 /organism="Onchocerca volvulus" /strain="Kumba, Cameroon" /db_xref="taxon:6282" /clone="SMML3C0758" /clone_11b="Onchocerca volvulus molting L3 larva cDNA (SL96MLN-OvML3)" /dev_stage="molting L3" /lab_host="XLI-Blue MRF" /note="Vector: lambda Uni-Zap XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (forest strain). The L3 were cultured in 20% FCS in IMDM+NCTC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volvulus molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (ML3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by Rnase H and DNA pol I. The library was constructed in the lambda Uni-Zap XR vector and has 1 x 10E6 independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams The library is available from Dr. Sara Lustigman (email: slustigman@bc.org)."
BASE COUNT	126 a 126 c 93 g 144 t 12 others
ORIGIN	
Query Match	15.1%; Score 269.4; DB 36; Length 466;
Best Local Similarity	75.3%; Pred. No. 3e-44;
Matches 327; Conservative	0; Mismatches 107; Indels 0; Gaps 0;
160	ttctgaatgaatgtgtgtcttattgtcttctgtactactatgtatgtcttattc 219
33	tactcatattatcatagcttgcgttaataagcgtattgtattatcaccgttaatcc 92
220	gattccggttcgaactgtgtcgaaggttagcagaatgaattgaattgacgaacttcaat 279
93	tattccgggtgtatgaacgggtgtgaaggtgtaacacaaattgaattgatgtgtccaaactcaatt 152
280	aacaactcaatttaatacagtaatgacatctcgaagagacatgttctatgtaaagtcctta 339
153	tactgtcaactttattatctcgtaattcatttgaagacactntntnctgtaagaccttata 212
340	tgatcaagaaggttcgcgttaatgaatgaaggttagcagtcgaagttgcggaaattcaactcc 399
213	cgatcagcagagatcccaaatgaaatgaagagtgtagcgtacgtagtgcgaatgacacttcc 272
400	attgatcatgcaaatgttggtgcgtaacgatactcttgataatccagttgattttgttaac 459
273	gattgatattcgttaattgagcagctgacacgttgcgttaanttcacactggtattgtgtcagc 332

QY	460	aaacaatgttgcattcgttccatccatttgcacaaagtgcagtcgcatcg	519
	1		
	2		
	3		
	4		
	5		
	6		
	7		
	8		
	9		
	10		
	11		
	12		
	13		
	14		
	15		
	16		
	17		
	18		
	19		
	20		
	21		
	22		
	23		
	24		
	25		
	26		
	27		
	28		
	29		
	30		
	31		
	32		
	33		
	34		
	35		
	36		
	37		
	38		
	39		
	40		
	41		
	42		
	43		
	44		
	45		
	46		
	47		
	48		
	49		
	50		
	51		
	52		
	53		
	54		
	55		
	56		
	57		
	58		
	59		
	60		
	61		
	62		
	63		
	64		
	65		
	66		
	67		
	68		
	69		
	70		
	71		
	72		
	73		
	74		
	75		
	76		
	77		
	78		
	79		
	80		
	81		
	82		
	83		
	84		
	85		
	86		
	87		
	88		
	89		
	90		
	91		
	92		
	93		
	94		
	95		
	96		
	97		
	98		
	99		
	100		
	101		
	102		
	103		
	104		
	105		
	106		
	107		
	108		
	109		
	110		
	111		
	112		
	113		
	114		
	115		
	116		
	117		
	118		
	119		
	120		
	121		
	122		
	123		
	124		
	125		
	126		
	127		
	128		
	129		

Email: genome@smith.edu
Seq primer: pBluescript SK.
Location/Qualifiers

FEATURES
source

```
1..395
/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMML3C0318"
/clone_id="Onchocerca volvulus molting L3 larva cDNA
(SL196MLM-Ovml3)"
/def_stage="molting L3"
/lab_host="Xrl-Blue MRF"
/note="Vector: Lambda Uni-Zap XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dt) followed by RNase H
and DNA pol I. The library was constructed in the lambda
Uni-Zap XR vector and has 1 x 10E6 independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustig@nybc.org)."
```

BASE COUNT 118 a 76 c 88 g 112 t 1 others
ORIGIN

```
Query Match 14.6%; Score 260.4; DB 36; Length 395;
Best Local Similarity 84.0%; Pred. No. 1.9e-42;
Matches 305; Conservative 0; Mismatches 57; Indels 1; Gaps 1;

QY 670 attgcatacttggtcagcagttatcataatgacatgcgattcctgaaccgtga 729
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 14 ATTTGCTACCAATGGTCA-CCAGTGTACCAATAATGACATGTGATCTGAACAGTTGA 72

QY 730 tacttcgcgcggtgtcctcctgctgttgatgtaagtgatgtaactgtga 789
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 73 TACATTCTGCTGTTGACTCATGCTTCTGATGATGTAACGGTGATAAGTGGA 132

QY 790 aattcaaatgctgagtgtgctctgtataaatatttgctaaataattggaatcc 849
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133 ATTATTAATCTGATGTTGCTTGACAAATATTGCTCAACAACTTGGATATCC 192

QY 850 aacagatttaagctgccaagaagctcacgtatataatgcygatacagct 909
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 193 AACTGATTTAATGCTGACAGAAGCTCACGCTACAAATATGCGATCATCACAACT 252

QY 910 tttaataagcagatcagatattacattaaagaacaataagcgaatgtgtgacc 969
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 253 GTTCTACCAATGTAATTAATTAATTAAGAACCAACAGTAGTGTGACGACC 312

QY 970 acaatgttcaagaccagaagctcgagctgttaaacaggtgtgcgcgagcaaac 1029
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 313 ACAATGTGCAGAACCAAGATTGGAGCATTCGNTCAGCAATGATGCTGCACGC 372

QY 1030 tgc 1032
    |||
Db 373 TGC 375
```

Search completed: April 16, 2000, 03:08:58
Job time: 6182 sec

Query Match	15.1%;	Score 269.2;	DB 39;	Length 354;
Best Local Similarity	85.0%;	Pred. No. 3.3e-44;		
Matches 301; Conservative	0;	Mismatches 53;	Indels 0;	Gaps 0;

OY	804	atggatgtgctcttgataataattgttgcnaaatcaatttggataataatccaaagatttaatg	863
Db	1	ATGTTGTGCTCTCGACAATATTTGGCTAAACAATTTGGAAATATCCAAAGGATCTTAATG	60
OY	864	ctggccaagaagctcacgataacaatatgagatcgatcacagaatttctatcaatgac	923
Db	61	CTGGTCAGAAACGGCATGTGTCAAAATATGCCGATGATCGCAACTTTTCTATWCATGTC	120
OY	924	agatcagiatcacctlaaagaacaatagcgaattgtctgcacacaatttgaagac	983
Db	121	AGATCAGATATCTCTTTAAAGAACCAACACAGTGAATGTCTCGCCACATTTTAGAGC	180
OY	984	cacaagaattcggagctgttaaacagaatgctgcgcagcaaaaacctgtcgaagctgac	1043
Db	181	CGCAAGGATTTTGGACCTTTTAAACGGGAGAGTGC CGCAGTAAACTCGAGACGCTGTC	240
OY	1044	aacttcgcttactcaagaaaagatctgtcagaacccggaaatatcatatgtatgaactg	1103
Db	241	AGCTTCGCTTGCTCAGAAAAGGCTTGGGGAACCGAAGAACGCTCTCGACAGTGAGACTG	300
OY	1104	ataccaacaccttgaattgaattagcgaatgataatcaagcttgcagattgattac	1157
Db	301	ATATCAGGCGCCTTGAATCATCTCGAAGCAATCAAGCTCTACCACTTGACTTGC	354

RESULT	14
AA618952	
LOCUS	459 bp mRNA
DEFINITION	SMOV3MCRA20ZSK Onchocerca volvulus molting L3 larva cDNA (SL96M7W-OvmLJ3) Onchocerca volvulus cDNA SMLJ3C02020 5', mRNA
ACCESSION	AA618952
VERSION	AA618952.1 GI:2522828
KEYWORDS	EST.
SOURCE	Onchocerca volvulus
ORGANISM	Onchocerca volvulus

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1. (bases 1 to 459)
Williams, S.A., Litotze-Waniewski, M., Laney, S. and Lustigman, S.
Genes expressed in molting L3 larvae of *Onchocerca volvulus*
Unpublished (1997)
On Sep 12, 1996 this sequence version replaced gi:1405277.

Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel.: 413/5851836

FEATURES	Location/Qualifiers
source	1. .459

```

/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMML3C02020"
/clone_1lb="Onchocerca volvulus molting L3 larva cDNA
(S156MWt-Ovmlt3)"
/dep_stage="molting L3"
/lab_host="X11-Blue MRF"
/notes="Vector: lambda Uni-ZAP XR; Site.1: Eco RI; Site.2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by

```

day 5 in culture mRNA was isolated from approximately 6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the Lambda Uni-Zap XR vector and has 1 x 10⁶ independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustigman@nbc.org).

Query Match	14.9%	Score 264.6	DB 36	Length 459	1
Best Local Similarity	77.8%	Pred. No. 2.7e+43			
Matches 318; Conservative	0	Mismatches 91	Indels 0	Gaps 0	

QY	160	ttctgaagaatgtagtcctcgtctatctgtcttcgtacaaacttaatgcatgctctatctc	213
Db	25	TACGATATATTTCATGCTTCGGTTGAATTAAGCGATTCTCTATTATTAATCACTCGTTAAATGC	84
QY	220	gattccggtgtgacaatggtgtgcgaagtgagccagaacttgaaatggtgaccaactcaat	279
Db	85	TATTCCGGTGTATTAACGGGTGTAGAAGGTGAACCGAATTTGAATGTGTGTCACACATCAAT	144
QY	280	aacatcgaattttaatacaacgtaatgcatcgcgaagaagacatgcttaatgtaaaagcttta	339
Db	145	TACGTGTAACCTTATATACCTGTAATTCATTTGGAAGGACGNTGTACGTGAAGGCTATATA	204
QY	340	tgatcagaagaagttgcgcgtatgataagaagtgagacgtaacgtttgcggaaattcacttc	399
Db	205	CGATCAGACGACGAGATGCCGAATATGTAGAAGTGGACGTCAGAGTATGACCCGGAAATTAACCTTCC	264
QY	400	attgattcattgcaaatgltgcgcgtacacgactctgaaatccacgltgtaattcttgtaac	459
Db	265	GTTTGTATCTCTGTATATGTATACACGTAACACGTCGTGGTTAAATCCAGTGGGTATTTTGTGAC	324
QY	460	aacacacgtgtgcatcttcglttcatcactaatcttgttaccaaaagttgacgtgcatatcg	519
Db	325	ATCAGTTGTTGTATCATCTTCATTCATTCATTCATTTGTTGTGACAAAGTTGATGACACATATCG	384
QY	520	agtaacaatgctttacataggaacgtgataaacaagttagtagcacagattc	568
Db	385	AATACATGCTTTTACATGTAAGACGTGTAAAGACAGTTAAGGCTCAACTT	433

RESULT	15
AA625010	
LOCUS	
DEFINITION	AA625010 395 bp mRNA EST 12-NOV-1997 SMOY3MCA318K Onchocerca volvulus molting L3 larva cDNA (SL36MTM-OVMJ5) Onchocerca volvulus cDNA clone SmtL3CO318 5', mRNA sequence.

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 395)	Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.	Genes expressed in molting L3 larvae of <i>Onchocerca volvulus</i>	Unpublished (1997)	On May 9, 1995 this sequence version replaced gi:802391.
	Contact: Steven A. Williams			
	Molecular Parasitology			
	Smith College Department of Biological Sciences			
	Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA			
	Tel: 4135853826			
	Fax: 4135853786			

OM of: US-09-323-427-2 to: A_Geneseq_36:* out_format : pfs
 Date: Apr 16, 2000 4:41 AM

About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame_r2p,model -DEV=xlp
 -Q/cgna2_1/uspro_pool/us09323427/runat_14042000_170513_19878/app_query.fasta.1
 -DB=A_Geneseq_36 -QFMT=fastlan -SUFFIX=backtrans.rag
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -CGAPOP=4.500 -CGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -CGAPOP=6.000 -RGAPEXT=7.000 -XGAPOP=10.000
 -YGAPEXT=0.500 -DELCP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blomsum62 -TRANS=human40.ccd -LIST=45 -DOCALIGN=200
 -THH_SCORE=pcr -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
 -MILEN=0 -MAXLEN=100000 -USER=US09323427 -NCPU=6 -ICPU=3
 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-323-427-2
 Query length: 1739
 Database: A_Geneseq_36:*
 Database sequences: 188963
 Database length: 23686106
 Search time (sec): 139.200000

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
A_Geneseq_36:R20112	-	100.00	154.90	0.0402	935	1 AE-III (peptidylhydroxylase)
A_Geneseq_36:R20112	-	100.00	140.65	0.0325	1594	1 Sequence encoded by the 2nd r
A_Geneseq_36:R20112	-	100.00	160.18	0.0755	332	1 Bovine neurotrophic biglycan. F
A_Geneseq_36:R20112	-	95.50	156.98	0.0779	369	1 Rat neurotrophic biglycan. Prot
A_Geneseq_36:R20112	-	94.50	155.03	0.1001	369	1 Human neurotrophic biglycan. P
A_Geneseq_36:R20112	-	94.50	147.65	0.1373	693	1 Expression plasmid pUCP1799.9
A_Geneseq_36:R20112	-	94.50	144.92	0.1544	875	1 C-terminal prepro-C-terminal al
A_Geneseq_36:R20112	-	91.50	133.10	0.3881	1226	1 Yeast transcription regulatory
A_Geneseq_36:R20112	-	91.00	124.64	0.6605	2756	1 Biosynthetic enzyme of iocospa
A_Geneseq_36:R20112	-	90.50	147.24	0.2726	368	1 Sequence of human bone proteog
A_Geneseq_36:R20112	-	90.50	131.07	0.5451	1464	1 Feline infectious peritonitis
A_Geneseq_36:R20112	-	89.50	145.09	0.3532	374	1 E. coli colitoxin or glucose tra
A_Geneseq_36:R20112	-	89.00	119.43	1.15	3079	1 GAP protein Iraz. Blocking Ras
A_Geneseq_36:R20112	-	88.00	124.89	1.08	1634	1 Enzyme involved in eicosapenta
A_Geneseq_36:R20112	-	88.00	118.77	1.40	2756	1 S. putrefaciens EPO biosynthes
A_Geneseq_36:R20112	-	87.00	131.66	0.9537	776	1 Lambda gt10ch2 encoded C-termi
A_Geneseq_36:R20112	-	87.00	130.37	1.01	866	1 Lambda gt10ch21 encoded C-term
A_Geneseq_36:R20112	-	87.00	128.99	1.07	974	1 Lambda gt10ch21 encoded C-term
A_Geneseq_36:R20112	-	87.00	115.67	1.89	3038	1 Aspergillus terreus tritol poly
A_Geneseq_36:R20112	-	86.50	128.38	1.19	944	1 Protein encoded by sequence whi
A_Geneseq_36:R20112	-	86.50	128.38	1.19	944	1 Type B alpha-amidating enzyme.
A_Geneseq_36:R20112	-	86.00	124.92	1.50	1167	1 Bacillus thuringiensis PSI67P
A_Geneseq_36:R20112	-	86.00	124.92	1.50	1167	1 Nematode toxin 167P protein. E
A_Geneseq_36:R20112	-	86.00	124.92	1.50	1168	1 Nematode toxin 167P protein. E
A_Geneseq_36:R20112	-	85.50	132.54	1.18	560	1 Segment of desmosomal cadherin.
A_Geneseq_36:R20112	-	85.00	125.10	1.76	973	1 C-terminal amidation enzyme. cd
A_Geneseq_36:R20112	-	85.00	113.10	2.95	2710	1 C. difficile toxin A. Fusion p
A_Geneseq_36:R20112	-	85.00	110.73	2.95	553	1 Clostridium difficile toxin A.
A_Geneseq_36:R20112	-	84.50	133.06	2.28	980	1 Human CD45 for use in T lympho
A_Geneseq_36:R20112	-	83.50	128.78	1.94	553	1 Bovine peptidyl-glycine alpha-d
A_Geneseq_36:R20112	-	83.50	128.78	1.94	553	1 Newcastle disease virus glycop
A_Geneseq_36:R20112	-	83.50	123.10	2.47	898	1 Haematopoietin receptor Hu-B1.2
A_Geneseq_36:R20112	-	83.50	122.97	2.48	908	1 Haematopoietin receptor Hu-B1.2
A_Geneseq_36:R20112	-	83.50	122.32	2.55	960	1 Haematopoietin receptor Hu-B1.2
A_Geneseq_36:R20112	-	83.50	120.42	2.77	1129	1 Oat phytochrome A apoprotein.
A_Geneseq_36:R20112	-	83.50	111.63	4.04	2391	1 Cadamoyl-phosphate-synthetase
A_Geneseq_36:R20112	-	83.50	111.25	4.10	2471	1 Candida albicans histidine kin
A_Geneseq_36:R20112	-	83.00	122.54	1.79	369	1 Mouse protease-activated recept
A_Geneseq_36:R20112	-	83.00	126.77	2.29	604	1 Human secreted protein clone b

A_Geneseq_36:R20112 - 83.00 122.73 2.73 853 ! Protein encoded by sequence
 A_Geneseq_36:R20112 - 83.00 122.73 2.73 853 ! Rat C-terminal amidating enz
 A_Geneseq_36:R20112 - 83.00 122.73 2.73 853 ! Type A alpha-amidating enzym
 A_Geneseq_36:R20112 - 83.00 118.23 3.31 1252 ! B. sphaerics SLP. Host cel

seq_name: A_Geneseq_36:R20112

seq_documentation_block:

ID R20112 standard; Protein: 935 AA.
 AC R20112;
 DT 06-APR-1992 (first entry)
 DE AE-III (peptidylhydroxylase N-C lyase precursor).
 KW Amidation; PHL.
 OS Xenopus laevis.
 FH Key
 FT Location/Qualifiers
 FT 1..935
 FT /label= AE-III
 FT /note="Including PAM and PHL domains"
 FT 383..935
 FT /tag= b
 FT /product= PHL
 region
 EP-465404-A.
 PD 08-JAN-1992.
 PF 27-MAY-1991. 810399.
 PR 01-JUN-1990; JP-141678.
 PR 10-NOV-1990; JP-210535.
 PR 30-NOV-1990; JP-329911.
 PA (CIBA) CIBA GEIGY AG.
 PI Iwasaki Y, Shimoi H, Suzuki K, Ghisalba D, Nishikawa Y;
 PI Kawahara T, Kangawa K;
 DR WPT: 92-010570/02.
 DR N-PSDB: Q20269.

PT Novel DNA encoding peptidyl hydroxylase N-C lyase (PHL) -
 PT used to prepare PHL which can be used in the amidation of
 PT peptide(s) e.g. human calcitonin.
 PS Claim 4; Page 18; 28pp. English.
 CC The sequence was deduced from a cDNA insert from PAE-III-202-4
 CC (EHEM BP-1172). The vector serves as a source for a DNA fragment
 CC encoding PHL for the construction of an expression vector for the
 CC prepun. of recombinant PHL. The protein may be purified to
 CC comprise only residues 383-706 or 383-713. The PHL catalyses the
 CC reaction: R-Glyox -> R-NH2. It can be used to produce peptides with
 CC amidated C-terminal, e.g. calcitonin, growth hormone, LH-RH.
 SQ Sequence 935 AA;

alignment_scores:
 Quality: 100.00 Length: 391
 Ratio: 0.541 Gaps: 18
 Percent Similarity: 47.315 Percent Identity: 19.437

alignment_block:
 US-09-323-427-2/rev x R20112 ..
 Align seg 1/1 to: R20112 from: 1 to: 935

1517 TGTGACCACTTCAATTAATCAATTTAATCAATGATGATTCGA 1468
 ||| |||||:|||||:|||||: ||| ||| |||
 526 CysGlnProThrAspValAlaValAspProIleThrGlyAsnPhelheya 542
 1467 AGACACTGTTTATGGAAGGCTTTATGATCAAGAAGTCCCGTAATG 1418
 :|||
 542 IAlAspGlyTyr..... 546
 1411 ATGAAGGTGAGTCAAGTTCGGGATTCCTTCAATTCATTCATTCG 1368
 547Cys 547
 1367 ATATGCGCGTACAGATCTCTGAATCCAGCGTATTTGTA..... 1323
 ||| :|||:|||||:|||||:|||||: |||
 548 Asn...SerArgIleMetGlnPheSerProAsnGlyMetPheIleMet 563
 1323 1323

```

563 ntprglygluthtserasnvalproargproglylnpheargi 580
1322 .....ACAAACAGTGTGTGATTCGTTTCATCCATATGTTTACC 1281
580 lepronihslerleuthmetvalproasprngllyglinleucysvalala 596
1280 AAAGTTGATCGTCATATCGATGATCGTTTACATGAGAACGTATTA 1231
597 asparngluasnngly...Argillelncysphehisalagluthrglyas 612
1230 AACAGTTGATCGACAGTATGAGTATCGAATC...ACAAGTGTTCCT 1184
612 phevallys...glinlelynsiglnlurpheclyarggluvalrhea 628
1183 AAACTCMAATGTCCCGATGCCAGATATGCCCTTATGAATTTTGATGCT 1134
628 lavalserlyralpro.....gly 634
1133 GGACCAACGGGTCAACAGTTCATTTGCTATCATGTCACCCAGTTTA 1084
635 gly.....valleutyralavalasnnglylpsprotyrty 646
1083 TCATTAATGACATGCGATTCGAAACCGTTGATCTTCTGCGCGTTG 1034
646 T.....glytyrseralapro 652
1033 TCCATTCCTGCTTGTGATGATGCTAAGCGTGATGCTGGAATTTCTA 984
652 alngllylphmetleuasnpheserasnnglyasp..... 663
983 AATCTGATGATGCTGCTTGTATTAATATTG.....CTAATAATT 940
664 .....lleuasnphrheleproalarglysasnph 675
939 GGAATATCCAAAGATTTATGCGTGGCCAAAGACCTCAACCTAATTAAT 890
675 easpmertronihsasprlealalaaspaspelylthrylvalyag 692
889 ATGCGATGATGATGATGCTTGTCTCAATGCGACATGCT..... 849
692 lyaspralnhislaasnalaalvaltrpyspneserproserlyaslaglu 708
848 ..ATTACCATTAAGAACCAAT.....ACGAAATGTTTCGACC 811
709 hisargservallyslaslaclylegluvalglulhethrgluth 725
810 ACAATGTTCAAGACCAAGATTCGAGCTGTTAAACAGCGTGGCCG 761
725 rglulrpehgluthhisileargserarprolysthrasnlgusery 742
760 CAGCAAAACCTGCTGACGCTGCCCACTTCGTTACTCAAGAAAGATCT 711
742 alglulysglnthrglulysglnglulys.....glulysasnser 756
710 GCA.....GAACGGAAGATATCATTT...GATGTAGAAC 679
757 Alaglyvalserthrglulysglinsnvalvalglnglulheasnal 773
678 TGATATCAACACCTTGA.....ATTAGGATGATTAATCAACT 638
773 aglyvalrprothrglulysglinsnvalvalglnglulseralag 790
637 TGCCAGTTGATTTACGTCACCGTGCACCTTCGCAACATTAATGACAACT 588
790 lyvalserthrglulysglinsnvalvalglnglulseralagl 806
587 GTATATCTGCTGACGTAACAATGATTCGATGTCACCATTTGCTT 538
807 valserthrglulysglinsnvalvalglnglulseralagl 823
537 CTCATGTTTATGCTTAAAGCATTCATGATTT...GCTGCCGTCATTA 491

```

```

823 lserphevalleullelthleuullelleleproillealvalleu 840
490 TTACCATTTGCTTTAAATTTGCT 468
840 lealalleallellellele 847

```

seq_name: A_Geneseq_36.P81184

seq_documentation_block:
ID P81184 standard; Protein: 1594 AA.
AC P81184;
DE 29-OCT-1992 (first entry)
DE Sequence encoded by the 2nd reading frame of the peplomic gene
DE of FIV strain 79-1146.
KW Vaccine; peplomic protein; E2 gene.
OS Feline infectious peritonitis virus.
PN EP-264979-A.
PD 27-APR-1988.
PF 01-SEP-1987; 201657.
PR 05-SEP-1986; NL-002244.
PA (DUIIN) DUPHAR INT RES BV.
PA De Groet RJ, Span WJM, Van Der Zeijst BAM;
DR WPI; 88-114147/17.
DR N-PSDB: N81533.
PT Gene for feline infectious peritonitis virus - and gene prod.
PS useful as antigenic protein for vaccine
PS disclosure: Fig 1, 13pp; English.
CC CDNA was prepd. from FIV strain 79-1146. N81533 gives the sequence
CC of the peplomic gene in three reading frames. The top reading
CC frame is an open reading frame of 4356 nucleotides and has a coding
CC capacity for a precursor polypeptide having a mol. wt. of 160,470
CC (1452 AAs). The beginning and the end of the E2 gene are indicated
CC in the FT of N81533. The first 18 N-terminal AAs have a strong
CC hydrophobic character and presumably comprise a cleavable signal
CC peptide. The extreme carboxy terminal part comprises a region of 20
CC hydrophobic AAs, which presumably serves as a transmembrane anchor.
CC The FIV peplomic protein has 35 potential glycosylation sites,
CC of which 22 are in the N-terminal part (pos. 1-790) which corresponds
CC to the S-part of the IBV E2 (see P81183). N.B. IBV = infectious
CC bronchitis virus. "X" in the AA sequence denotes the translation
CC of a stop codon.
SQ Sequence 1594 AAs:

alignment_scores: Length: 581
 Quality: 100.00
 Ratio: 0.376
Percent Similarity: 45.783 Percent Identity: 18.072

alignment_block:
US-09-323-427-2/rev x P81184 ..

Align seg 1/1 to: P81184 from: 1 to: 1594

```

1586 ACTACACTTATTCATGTTCTTATTCGATTCGCGTTGACAAATGTCGCA 1537
      ::::::::::::::::::::|::|::|::|
51 SerThrIleMetAlaLeuSerHis.....AsnSerValle 62
1536 AGGTGAGCCGGAATGATGTGACCAACTTCATTAACAATTAATTTTA 1487
      ||::|::|::|::|::|::|::|::|::|::|::|
62 uasprThrProlySerPheLysCysArgCysAsnAsnAlaLeuIle-ValAsn 78
1486 ATACAGTAAATGCAATTCGAGAGACATTTATGTAAGGCTTTATGAT 1437
      ::|::|::|::|::|::|::|::|::|::|::|::|
79 LeuLysGluLysGluLeuAsnGluMetVal.....ValGlyLeuLe 92
1436 CAAGAAGTTCCTGAATGATGAAGTGAGAGTCAAGTTCGCGGAATTTTC 1387
      ::|::|::|::|::|::|::|::|::|::|::|::|
92 uArgLysGlyLysLeuLeuIleArgAsnAsnGlyLysLeuLeuAsnPhg 109
1386 ACTTCATTTGATTCATGCAATGTTGCGGTACACAGATCTGATCCAC 1337
      |||::|::|::|::|::|::|::|::|::|::|::|
109 LyAsnHisLeu.....Valasn 114

```


PS Claim 3; Fig 8: 60pp: English.
 CC Bovine biglycan (R87953) is a chondroitin sulphate proteoglycan with
 CC neurotrophic activity for brain neurons. It can be used to enhance
 CC the survival and maintain the structure and function of CNS neurons
 CC during normal ageing as well as after pathological and/or traumatic
 CC nervous system damage. It can also be used to restore function
 CC following nervous system lesions and degenerative diseases, and to
 CC improve learning efficiency and memory in the elderly and in patients
 CC with dementia.
 SQ Sequence 332 AA:

alignment_scores: Length: 299
 Quality: 96.50 Gaps: 13
 Ratio: 0.798
 Percent Similarity: 40.468 Percent Identity: 20.401

alignment_block:
 US-09-323-427-2/rev x R87953 ..

Align seg 1/1 to: R87953 from: 1 to: 332

```

775 AAACAGGTGTCGCCGACGAAACCTGCTGACGCTGCGCAACTTCGTTTA 726
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
50 Lysglutileserproaspthrthreuleuaspheuglnasnaspsil 66
725 CTCA.....AGAAAGAT.....713
    |||
66 esercluleuarglyaspaspheylsleuglnhisleutyrylall 83
    |||
712 .....CTGAGAACCGGAGAAATATCATTCATGTACGAACTGATATCACA 668
    |||
83 euvalleuvalnasnlyslleserlyllehisglulysalapheser 99
    |||
667 CCCTTGAATTTAGCGATGATTAATCAACTTTGCCACTGATTTACTGTCAC 618
    |||
100 Proleuarg.....Lysleuglnlyslsleutyrylleserly 111
    |||
617 CGTGCACTTCGACATATGACAACTGTAATCTTGCTGCTGACGTACA 568
    |||
111 sasnhisleuvalglulileproproasn.....120
    |||
567 AATGGAATGTCATGTCACCATTTGGCTTCTCAATGTTATGGGTTAA 518
    |||
120 .....120
517 GCATTCATTCATTCCTGCCGTCTATTACATTTGCTTTAATTTGCT 468
    |||
121 .....LeuproserSerleu.....Va 126
    |||
467 CCAATATCAGAGCATATAAATATGTTAGA.....ATCA 433
    |||
126 lglleuargllehisaspasnargllearglyvalprolysglyvalp 143
    |||
432 TCGAAGCAATA.....422
    |||
143 heserlyleuargasmetasnlgulileglumetglylyasnproleu 159
    |||
421 .....ATAAACTGCCATA 408
    |||
160 Gluasnseryglyphegluproglyalaphespglyleuylsleuasnly 176
    |||
407 TATATTCGTTTCTTCTATCATCCTTCTAATACTAATTTTATGCTAACA 358
    |||
176 rleuarglileserglualalys.....leuthrg 186
    |||
357 AATATATGATGATGAGCAATATTAATTAATACATAAATGATATTTTC 308
    |||
186 lytle.....Prolysaspleuarglulthreuleasnlglyleuhs 199
    |||
307 ATCAAAACTTCTTCTATCGCTTTTATGCTTTCGAAAAGTTTATTCATTA 258
    |||
200 leuaspHisasnlyslleglnalalalegluleugluaspheuleuargly 216

```

```

257 TTCAGTATCTTTATATGACATATTTGTAATATGTTTCATCATTAAGCC 208
    |||
216 rserlyslleutyrylrgleu.....Glyleuglyh 226
    |||
207 ATGCAATAGTTTCTGTTGTTATATCA...TCATTATCACTTGCTCT... 164
    |||
226 lasnclnleargmetllegluasnglyserleuserpheuargleu 242
    |||
164 .....164
243 leuarggluleuhsleuaspnasnlyslseuserargvalproalagl 259
    |||
163 .....ATTTATTTCTAACAGTTTATCATTTGATGATATA 130
    |||
259 yleuproaspheuleuylsleuglnvallytyrleuhsisThrasmnsnI 276
    |||
129 TCACAAATTTATCTTG.....TATGCCCAATTTTATAGGCGATC 89
    |||
276 lenthlyslvalglyvalasnaspheycysprovalglypheglyval 291
    |||

```

seq_name: A_Geneseq_36:R87951

seq_documentation_block:

```

ID R87951 standard; Protein; 369 AA.
AC R87951;
DE 20-MAR-1996 (first entry)
DE Rat neurotrophic biglycan.
DE Biglycan, proteoglycan; chondroitin sulphate; neuron protection;
KW neurotrophic; central nervous system; CNS; memory loss; dementia;
KW learning.
OS Rattus sp.
FH Key Location/Qualifiers
FT peptide 1..37
FT /label- Sig-peptide
FT region 44..60
FT /label- Hypervariable_region
FT WO9530432-A1.
FT 16-NOV-1995.
PD 09-MAY-1994; E01479.
PR 09-MAY-1994; WO-E01479.
PA (BOE) BOEHRINGER MANNHEIM GMBH.
PI Hasenoehtl R, Huston J, Junghans U, Kappler J, Koops A;
PI Mueller HW;
DR WPI; 95-403938/51.
DR N-PSDB; T08768.
PT Proteoglycan cpds., partic. chondroitin sulphate proteoglycan(s)
PT for maintain structural and function of the CNS and attenuating
PT memory deficit(s) in the elderly and patients with dementia
PS Claim 1; Page 44-45; 60pp: English.
CC Rat biglycan (R87951) is a chondroitin sulphate proteoglycan with
CC neurotrophic activity for brain neurons. Recombinant biglycan,
CC obid. by expression of encoding cDNA (T08768) in eukaryotic host
CC cells, can be used to enhance the survival and maintain the structure
CC and function of CNS neurons during normal ageing as well as after
CC pathological and/or traumatic nervous system damage. It can also
CC be used to restore function following nervous system lesions and
CC degenerative diseases, and to improve learning efficiency and memory
CC in the elderly and in patients with dementia.
SQ Sequence 369 AA:

```

alignment_scores: Length: 299
 Quality: 95.50 Gaps: 13
 Ratio: 0.789
 Percent Similarity: 40.468 Percent Identity: 20.401

alignment_block:

US-09-323-427-2/rev x R87951 ..

Align seg 1/1 to: R87951 from: 1 to: 369

775 AAACAGGTGTCGCCGACGAAACCTGCTGACGCTGCGCAACTTCGTTTA 726

```

87  LysGluIleSerProAspThrLeuLeuAspLeuGlnHisAsnAspI1 1030
725  CTCA.....AGAAAGAT.....713
103  eSerGluLeuArgLysAspAspPheLysGlyLeuGlnHisLeuTyAlaL 1200
712  ....CTGCAGACCGGAGATATATCATCTTATGATGACGACTGATATCA 668
120  euValLeuValAsnAsnLysIleSerLysIleHisGluLysAlaPheSer 136
667  CCCTTGAATTATGACGATGATATCAAGCTTTGCCACTGTGATTACGTCAC 618
137  ProLeuArg.....LysLeuGlnLysLeuTyIleSerIy 148
617  CGTGCACCTTTCGCAACATPAATGGACACCTGTAAATCTTGCTGCAGTACA 568
148  sAsnHisLeuValGluIleProProAsn.....157
567  AAATGGAAATGTGCATGTCCACATTTGGCTTCGATCATGTTATGGTTAA 518
157  .....157
517  GCATTTGCATTTGATTCGTCGCCGTCATTTATACCATTTCTGTTAAATTGCT 468
158  .....LeuProSerSerLeu.....Va 163
467  CCAATATCAGAAAGCATATAAAATATGTTAGA.....ATCA 433
163  IGlLeuArgIleHisAspAsnArgIleArgLysValProLysGlyValP 180
432  TCGAAGCAATA.....422
180  heSerGlyLeuArgAsnMetAsnCysIleGluMetGlyLysAsnProLeu 196
421  .....ATAAACTGGCATA 408
197  GluAsnSerGlyPheGluProGlyAlaPheAspGlyLeuLysLeuAsnTy 213
407  TAATTTGGTTCTTCTTAATCATCCTCTTAATPAACATAATTTAGCTACAA 358
213  rLeuArgIleSerGluAlaLys.....LeuThrG 223
357  ATATATAGTATGTGGAATATTAATCTATATACAAATAAGCATATTTC 308
223  Lylle.....ProLysAspLeuProGluThrLeuAsnGluLeuHis 236
307  ATCAAAACTCTTCTATCGCTTTTATAGCTTCTGAAGAAGCTTATTCATTA 258
237  LeuAspHisAsnLysIleGlnAlaIleGluLeuGluAspLeuLeuArgTy 253
257  TTACAGTAATCTTTATATGCATACACTATGTAAATGTTGCATATTAGGCC 208
253  rSerLysLeuTyArgLeu.....GlyLeuGln 263
207  ATGAATAGTTCTGCTGTTATATCA...TCATATGCACACTTGCC... 164
263  iAsnGlnIleArgMetIleGluAsnGlySerLeuSerPheLeuProThr 279
164  .....164
280  LeuArgGluLeuHisLeuAspAsnAsnLysLeuSerArgValProAlaG 296
163  .....ATTATTTCTACAGTTTATGATTTGTGATATAATA 130
296  yLeuProAspLeuLysLeuLeuGlnValValTyLeuHisSerAsnAsnI 313
129  TCACAAATTATACCTTG.....TATGGCCAAATTTTATGGGATC 89
313  LeuThrLysValGlyIleAsnAspPheCysProMetCysGlyPheGlyVal 328
seq_name: A_Geneseq_36:R87952

```

```

seq_documentation_block:
AC R87952: standard; Protein; 369 AA.
AC R87952:
DT 20-MAR-1996 (first entry)
DE Human neurotrophic biglycan.
KW biglycan; proteoglycan; chondroitin sulphate; neuron protection;
KW neurotrophic; central nervous system; CNS; memory loss; dementia;
KW learning.
OS Homo sapiens.
PI key Location/Qualifiers
PI peptide 1..37
FT /label= Sig_peptide
FT region 44..60
FT /label= Hypervariable_region
MO530432-A1.
PD 16-NOV-1995.
PF 09-MAY-1994; E01479.
PR 09-MAY-1994; WO-E01479.
PA (BOE) BOEHRINGER MANNHEIM GMBH.
PI Hasenlechner R, Huston J, Jungmans U, Kappler J, Koops A;
PI HMEI.
PI WPI: 95-403938/51.
PT Proteoglycan cspds., partic. chondroitin sulphate proteoglycan(s) -
PT for maintain structure and function of the CNS and attenuating
PT memory deficit(s) in the elderly and patients with dementia
PS Claim 3; Fig 8; 60pp; English.
CC Human biglycan (R87952) is a chondroitin sulphate proteoglycan with
CC neurotrophic activity for brain neurons. It can be used to enhance
CC the survival and maintain the structure and function of CNS neurons
CC during normal ageing as well as after pathological and/or traumatic
CC nervous system damage. It can also be used to restore function
CC following nervous system lesions and degenerative diseases, and to
CC improve learning efficiency and memory in the elderly and in patients
CC with dementia.
SQ Sequence 369 AA;

```

[illegible]

```

158 .....LeuproserSerLeu.....Va 163
467 CCNAATCAGAAAGCATAAATATATGTAGA.....ATCA 433
163 |G|U|e|u|e|a|r|g|i|l|e|i|s|p|a|s|n|a|r|g|i|l|e|a|r|g|y|v|a|l|p|r|o|l|y|g|i|y|a|l|p 180
432 TCGAAGCAATA..... 422
180 hserGlyLeuArgAsnMetAsnCysIleGluMetGlyGlyAsnProLeu 196
421 .....ATTAACCTGCATA 408
197 GluAsnSerGlyPheGluProGlyAlaPheAspGlyLeuLysLeuAsnTyr 213
407 TATATTCGTTCTTCTTATCATCTCTTCAATAACAAATTGTAACAA 358
213 rLeuArgIleSerGlyAlaLys.....LeuIrrg 223
357 ATATATAGTATCTAGGAATATATCTGTAATACAAATAAGTATTTTC 308
223 TyIle.....ProLysAspLeuProGlyThrLeuAsnGluLeuHis 236
307 ATCAAAACTTCTTCTATTCGCTTTTATAGCTTCTGAAGAAGCTTATTCATA 258
237 LeuAspHisAsnLysIleGlnAlaIleGluLeuGluAspLeuLeuArgTy 253
257 TTCAGTATCTTTTATATGCAATCACTATGTAATGTTTCATCATTTAGGCC 208
253 rSerLysLeuTyArgLeu.....GlyLeuGlyH 263
207 ATCAATAGTTCGTTGTTGTTATATCA...TCATATTCACCTTGCT... 164
263 lAsnGlnIleArgMetIleGluAsnGlySerLeuSerPheLeuProThr 219
164 ..... 164
280 LeuArgGluLeuHisLeuAspAsnAsnLysLeuSerArgValProAlaG1 296
163 .....ATTATTCCTAACAGCTTATATCATTTGTGATATA 130
296 yLeuProAspLeuLysLeuGlnValIleTyIleuHisSerAsnAsnI 313
129 TCACAATTTATACCTTG.....TATTGCCAATTTTATGGGCATC 89
313 lerHisValAlaGlyValAsnAspPheCysProMetGlyPheGlyVal 328
seq_name: A_Geneseq_36:P94856
seq_documentation_block:
ID P94856 standard; protein; 693 AA.
AC P94856;
DE 27-JUN-1990 (first entry) BgIII gene product.
DE Expression plasmid pUCP1799
DE alpha-amidating; pAX799; alpha amide; ds.
KM Synthetic.
OS EP-299790-A.
PN 18-JAN-1989.
PD 15-JUL-1988; 306508.
PE 17-JUL-1987; JP-177184.
PR 05-DEC-1987; JP-306867.
PR (SUNR) Suntoory Ltd.
PA Ohsuye K, Kitano K, Tanaka S, Matsuo H, Mizuno K;
PI WPI: 89-011279/03.
DR N-PSDB: N94527.
PT Recombinant C terminal alpha amidating enzymes of Xenopus laevis -
PT and their precursors deoxyribonucleic acid encoding sequences.
PS Disclosure: 7pp: English.
PS The sequence encodes a derivative of the mature C-terminal alpha-
CC amidating enzyme from plasmid pXA457.
CC The plasmid was screened from an E.coli library using plasmid pXA457
CC to screen a larger library.
CC Although pXA799 is similar to pXA457 at the N-terminus, it has an area of
CC hydrophobic elements suggesting a membrane function.

```

```

CC See also N93060.
SQ Sequence 693 AA;
alignment_scores:
Quality: 94.50 Length: 232
Ratio: 0.945 Gaps: 12
Percent Similarity: 43.103 Percent Identity: 21.552
alignment_block:
US-09-323-427-2/rev x P94856 ..
Align seg 1/1 to: P94856 from: 1 to: 693
1517 TGTGACCACTTCATATACATCATTTTAAATACAGTATGCTTGA 1468
1491 CysGlnProThrAspValAlaValAspProIleThrGlnAsnPhePheVa 507
1467 AGGCATGTTTATGTGAAGGCTTATGATCAAGAGTTCGGTATG 1418
507 lAlaAspGlyTy..... 511
1417 ATGAAGTGAGCGTCAAGTTCGGGAATTCACCTTCATTTGATTCATGC 1368
512 .....Cys 512
1367 AATGTGGCCGTACACGATCTCTGAATCCAGCTGATTTTGTGA..... 1323
513 Asn...SerArgIleMetClnPheSerProAsnGlyMetPheIleMetG1 528
1323 ..... 1323
528 nTrpGlyGluGlnThrSerSerAsnLeuProArgProGlyGlnPheArgI 545
1322 .....ACAACTGTTGTCATTTGCTTTCATTCATTCATTTGTTAC 1281
545 lProHisSerLeuThrMetClnSerAspGlnGlyGlnLeuGlyValAla 561
1280 AAAGTATGTCGATATGAGTACATGCTTTTACATGGAAGCTGATA 1231
562 AspArgGlnAsnGly...ArgIleGlnCysPheHisAlaLysTrpGlyG1 577
1230 AACAGTATGTCACAGATTTGAGTATCTGAATC...ACAACGCTTTTC 1184
577 uPheValLys...GlnIleTySHISGlnGluPheGlyArgGluValPheA 593
1183 AAACTCAAATGTCCCGATGCCAGTATGCCGTTATGAATTTTGGATGT 1134
593 lValSerTyAlaPro.....Gly 599
1133 GGACCAACCGGTCACACAGTTCAATTTGCTATCATTTGTCACCCAGTTTA 1084
600 Gly.....ValLeuTyAlaValAsnGlyLysProTyTrp 611
1083 TCATAATAGTACATGGATTCGAAACGCTTGATCTTCTGCGCGGTG 1034
611 rGly.....AspSerThrProValGlnGlyPhe..... 620
1033 TCCATTCCTGCTTTGTCGATGATGTAACGGTATGTAAGTAATTCGA 984
621 .....MetLeuAsnPheSerAsnGlyAsp..... 628
983 AATGCTATGATGTGCTCTGATAATATTTG.....CTAATAATTT 940
629 .....IleLeuAspThrPheIleProAlaArgLysAsnPh 640
939 GGAATATCCACAGATTTAATGCTGCGCCAGAAGCTCACGATATAC 894
640 eGluMetProHisAspIleAlaValAlaGlyAspAspGlyThrValTy 655
seq_name: A_Geneseq_36:P94854
seq_documentation_block:

```



```

1417 ATGAAGTGGAGTCAGTCAAGTTGGCGGATTTCACTTCATTTGATTCATGC 1368
1367 AATGTTGGCGGTCAGTCAAGTTCTGATTCAGTCAAGTGGATTTTGTGA.... 1323
551 .....Cys 551
552 Asn...SerArg11IleMetClnPheSerProAsnGlyMetPheIleMetC1 567
1323 ..... 1323
567 nTPPGlyGlnInThrSerSerAsnLeuProArgProGlyGlnPheArg1 584
1322 .....ACAACAAGTCTTTCATTTGATTCATTCATTTGTTTACC 1281
584 leProHisSerLeuThrMetClnSerAspGlnGlyGlnLeuGlyValAla 600
1280 AAGTGTGATGTCATATCGATGATGATGATGATGATGATGATGATGATGAT 1231
601 AspArgGlnAsnGly...Arg11IleGlnCysPheHisAlaLeuThrGlyG1 616
1230 AACAGTTCAGTCAAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGT 1184
616 upheValLys...GlnIleLysHisGlnGlnPheGlyArgGlnValPheA 632
1183 AACCTCAATTCGTCGATGCGCAGTATGCGGTATGAAATTTTGCATGCT 1134
632 laValSerTyrAlaPro.....Gly 638
1133 GGACCAACCGGTCACACAGTTCATTTGATTCATTCATTCATTCATTCATTCAT 1084
639 Gly.....ValLeuTyrAlaValAsnGlyLysProTyrTyr 650
1083 TCATTAATGACATGCGATTCGAAACCGTTCGATTCGATTCGATTCGATTCGAT 1034
650 rGly.....AspSerThrProValGlnGlyPhe..... 659
1033 TCCATTCCTGCTTTCGATGATGATGATGATGATGATGATGATGATGATGAT 984
660 .....MetLeuAsnProSerAsnGlyAsp..... 667
983 AATGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 940
668 .....IleLeuAspThrPheIleProAlaArgLysAsnIle 679
939 GGAATATCCACAGATTTTAATGCTGCGCAAGAGCTCAGCTATAC 894
679 eGluMetSerProHisAspIleAlaIleAspLysAspArgLysThrValTyr 694

seq_name: A_Geneseq_36:W13825
seq_documentation_block:
ID W13825 standard; protein: 1226 AA.
AC W13825:
DT 04-JUN-1997 (first entry)
DE Yeast transcription regulatory factor SRB8.
KW SRB8; RNA polymerase II; holoenzyme; SWI/SNF.
OS Saccharomyces cerevisiae.
PN W09708301-A1.
PD 06-MAR-1997.
PF 28-AUG-1996: U14192.
PR 31-AUG-1995: U5-521872.
PR 11-OCT-1995: U5-540804.
PR 26-JAN-1996: U5-590399.
PA (WHEED ) WHITEHEAD INST BIOMEDICAL RES.
PI Chao DM, Koloske AJ, Thompson CM, Young RA.
PI WPI; 97-179258/16.
DR N-PSDB; T59908.
PT Purified RNA polymerase II holoenzyme - comprises RNA polymerase II
PT and one or more regulatory proteins; pref. suppressor of RNA
PT polymerase B proteins or SWI/SNF proteins
PT Claim 11: Fig 10a-b; 154pp: English.
PS Novel yeast SRB (suppressor of RNA polymerase B) proteins SRB4, SRB5,
CC

```

```

CC SRB6, SRB7, SRB8, SRB9, SRB10 and SRB11 (W13821-28) are transcription
CC regulatory factors that act as positive and negative regulators of
CC RNA polymerase II activity, and are components of the RNA polymerase
CC II holoenzyme. They were identified using methods designed to
CC identify transcription factors involved in RNA polymerase II
CC C-terminal domain (CTD) function. SRB8 and SRB9 appear to repress
CC CTD activity. Genomic clones (T59904-11) for the SRBs have been obt.
CC SRBs can be used to treat diseases resulting from alteration or
CC deletion of the SRB gene, pref. by gene transfer technology. They
CC can also be used in in vitro transcription of DNA and to identify
CC cpds. that modify gene transcription.
CC Sequence 1226 AA.
50

```

```

alignment_scores:      Length: 489
Quality: 91.50         Gaps: 24
Ratio: 0.393
Percent Similarity: 47.648  Percent Identity: 19.632

```

```

alignment_block:
US-09-323-427-2/rev x W13825

```

```

Align seg 1/1 to: W13825 from: 1 to: 1226

```

```

1267 CATATCGATGACATGCTTTTACATGAGAGCTGATTAACAGTTA..... 1223
279 TyrValAspGlnAsnProLeuThrMetHisLysIleIleGlnLeuIleLe 295
1222 .....GTGCACAGATGAGC.....TATCTGAATCAGAAC 1192
295 uTPSerIleHisProSerArgGlnPheAspHisTyrGlnLysSerAsnGln 312
1191 TGCTTTTCACAACTCAATTTGTCGCGATGCCAGTATGCCGTTATGAAATTT 1142
312 euValAlaLysLeuLeuLeuLeuArgIleAsn..... 322
1141 TGGATGCTGGACCAACCGGTCACAGCTTCAT.....TTGCT 1104
323 ...SerThrAspGlnAspLeuHisGlnPheGlnIleGlnAspAlaIleTr 338
1103 ATCATTTGTCACAGCAGTTATGATGATGATGATGATGATGATGATGATGATGAT 1054
338 pSerLeuValIlePheGlnLeuAlaLysAsn..... 347
1053 TGATACTTTCTGCGCGTTGTCATCTGCTTTCGATGATGATGATGATGATGATGAT 1004
348 .....PheSerAlaGlnLysArgValIleSer 356
1003 GTGATACTGTGGAATTTCTAATGCTGATGATGATGATGATGATGATGATGATGAT 954
357 TyrMetMetProSerLeuTyrArgLeuLeuAsnIleLeu..... 369
953 TTGCTAATTAATTTGGAATATCCACAGATTTAATGCTGCGCAAGAGC 904
370 .....IleThrTyrGlyIleIleLysValProThrTyrIleArgLysL 384
903 TCACGAT.....ACAATATGCGGATGAT 878
384 euIleSerSerGlyLeuLeuTyrLeuGlnAspSerAsnAspLysPheVal 400
877 CAC...AGCTTTTATCATGATGCGCAGATGATGATGATGATGATGATGATGATGAT 417
401 HisValGlnLeuIleAsnLeuLysIleSerProLeuMetLysSerG1 417
830 A.....ATAGCAATGTGTTGACCCACACATGTTAGAACACCA 793
417 nTyrAsnMetValIleuArgAsnValMetGluTyrAspValLysPheTyrG 434
792 AG.....GATTCGAGCTGTTAAACAGGCTGCGCCAGCAAAACCG 749
434 IuIlePheAsnPheAspGlnValIleGluIleThrGlnGlnIleLysMet 450
748 CTCGAGCTGCGCACTTCGTTTACATCAGAAAGATCTCAGAACCGG 699

```



```

451 ArgIleLeuSerAsn.....AspIleThrAsnLeuG 461
698 AATATCATGTGATGTACGAACTGATATCAACACCCCTTGAATATGACGTGA 649
461 nleuser.....LysThrProLeuSerIleLysIleM 47Z
648 TAATC.....AAGCTTGGCAGTTGATTTACGTCACC 617Z
472 etValAlaGluTrpTyrLeuSerHisLeuCySerGlyIleLeuSerS 488
616 GTGCAC.....TTCGCAACAATAT 597Z
489 ValAsnArgThrValLeuLeuLeuLysIlePheLysIlePheCysIleAspLe 505Z
596 GGACAAACCTGTAAATCTGCTGCAGTACACAAAATGGAATCGCATGTCACC 547Z
505 uGluValPheHisHisPhePheLysTrpIleGluPheIleValTyrHisG 522Z
546 ATTGGCGTTCCAATGTTTAAAGCTTTAAAGCATTTGCATGATTGCTGCCG 497Z
522 InleuLeuSerAsp.....IleGluSerLeuGluAlaLeuMetAsp 535Z
496 TCATATTATACCATTTCTGTAAATTTGCGCAAA...ATCAGAGAGCAT... 452Z
536 IleuLeuLeuGlyTrpGlnLysLeuLeuPheSerGlnPheIleAsnAspHisI 552Z
451 .....AAAAATAATGTTAGATCATCGAACCATATAATAAAA... 416Z
552 eleuPheThrTyrThrPheIlePheIleTyrLysValLeuLysGluL 569Z
415 .....GTGCCATATATATCTGTTTCTTCTTATCATCCTCTAATACCTAA 371Z
569 ysaAspValProAlaTyrAsnValThrSerPheMetCProPheTrpLys... 584Z
370 TTTTAGCTAACCAATATATATAGTAT.....GTAGCAAAATAATATCTAG 330Z
585 PhePheMetLysAsnPheProPheValLeuLysValAspAsnAspLeuXr 601Z
329 TAATACATAAGTATATTTTC.....ATCAAACTTCTTCTTA 292Z
601 gIleGluLeuGlnSerValTyrAsnAspGluLysLeuLysTrpGluLysL 618Z
291 TCGCTTTATAGTCTTCGAAAGCTTATCATTTATTCAGTATTC... 248Z
618 euLysAsnAspLysSerGluValLeuLysValTyrSerMetIleAsnAsn 634Z
248 ..... 248Z
635 SerAsnGlnAlaValGlyGlnThrTrpAsnPheProGluValPheGlnVal 651Z
247 .....TTTATATATGCATACTATTTGTAAGTCTTATCATCATTAAGGCC 208Z
651 lasnIleArgPheLeuHisAsnSerGluIleLeuAspThrAsnThrS 668Z
207 ATGAATATTTTCGTTTGTATTATTCATCATCATTAACAATGTCCTATTTTA 158Z
668 erLysGlnPheGlnLysAlaArgAsnAsnVal..... 678Z
157 TTCTAACAGTTTATCATTTGTGATTAATATCAACAATTTATACCTTTATTTG 108Z
679 .....MetLeuLeuIleAlaThrAsnLeuLysGlnTyrAsnLysPheMe 693Z
107 CCCAATTTTATGGGCATCTTCCCTATTCGTAAACAACTTACCTATTATTT 58Z
693 tSerIlePheLeuLysArgLysAspPheThr...AsnLysAsnLeuIleG 709Z
57 GCATTTATTCGCAATTA 41
709 InleuIleSerLeuLys 714
seq_name: A_Geneseq_36:R99462

```

seq_documentation_block:

ID R99462 standard; Protein; 2756 AA.

AC R99462;

DE 30-JAN-1997 (first entry)

DT Biosynthetic enzyme of icosaepentaenoic acid synthase.

KE Icosaepentaenoic acid synthase; EPA; drugs; agrochemicals;

KW foodstuffs; animal feed; lipid balance correction; antihypertensive;

KW antinflammatory; anticancer agent.

OS *Shewanella putrefaciens*.

PN W09621735-A1.

PD 18-JUL-1996.

PE 12-JAN-1996; J000030.

PR 13-JAN-1995; JP-004299.

PA (SAGA) SAGAMI CHEM RES CENTRE.

PI Kato S, Kondo K, Yamada A, Yazawa K;

DR WPI: 96-342288/34.

DR N-PSDB: T34137.

PT Production of icosaepentaenoic acid using transformed *E. coli* - uses

PT DNA coding for icosaepentaenoic acid synthase derived from *Shewanella*

PT strain

CS Claim 7; Page 83-94; 145pp: English.

CC The DNA sequence (T34137) which encodes the biosynthetic enzymes of

CC Icosaepentaenoic acid (EPA) can be used to transform *Escherichia coli*.

CC The DNA sequence allows efficient microbial production of EPA, which

CC is a raw material for drugs, agrochemicals, foods and animal

CC feedstuffs. EPA is also useful for lipid balance correction and as

CC an antihypertensive, antinflammatory and anticancer agent.

CC Sequence 2756 AA;

```
alignment_scores:
  quality: 91.00
  ratio: 0.389
  gaps: 23
Percent Similarity: 46.800
Percent Identity: 18.400
```

```
alignment_block:
US-09-323-427-2/rev x R99462 ..
```

Align seg 1/1 to: R99462 from: 1 to: 2756

```

1745 CGCGATCGCAGCAGCAACCAACCAACCAACCAACCAACCAACCAACCAACG 1695
      ||| ||| :::: ||| ||| ||| |||
827 ProhehlaIySalalIeaspaIaIalysIepheIlystIheSerIal 843
1695 AATATATACCCCATCAAGTGAGAGAAAGACAGAGCAATCTAGTT 1646
      :::: :::: :::: |||||
843 aleuTySerSaIasIaIethIyIyIy...LeuTyrgIuseThIraIaIaI 859
1645 TTCTAAAAATCGAATTACTAAATCTTCTGGAATGATCTGTTATT 1566
      :::: ||| ||| ||| ||| ||| ||| |||
859 ystIeIySaIasIaIerPheIySlyshs.....MetIeudIseIerVal 872
1595 GCTTTTGTGACTACATTTAGTCATGTCTTATTCGATTCGGTGAAGA 1546
      ||| ||| ||| ||| ||| ||| ||| |||
873 ArgPheThSerIeugIuIaIaIeT.....TyraSas 884
1545 TGGTGTGGAAGGTGAGCCAGCAAAATTGAATGTGACCACTTCAATA... 15000
      ||| ||| :::: ||| ||| ||| ||| ||| |||
884 pGIy.....AlaIarValaPheValGIuPheGIyProlysaSnIleuG 899
1499 .....ACA 1497
899 IuIySeIueValaIeGInGlyIhIeValaSnIhIuIaSnGIuValaIyStyr 915
1496 ATCAATTTTAATACAGTAAATCATTTGGAAGACATGTTATGTGAAG 1447
      ||| ||| ||| :::: ||| ||| ||| ||| ||| |||
916 IIsaerIleasIproIasIproIySaIaSpseIaIeugIuIeulys.. 931
1446 TCTTTATGATCAAGAAGGTGCCCTTAATGATGAAGGTGACCTGAATG 1397
      :::: ||| ||| ||| ||| ||| ||| |||
932 .....GlaIaIaIaIeIuIeIuA 938
1396 CC.....GGAATTCACCT.....CAATTGATTCATGCAAT 1365

```



```
931 ..... 931
874 AGCTTTTCATCAATGCCAGATTGACATTAACCATTAAAGAACCAAATAGC 825
      |||||||::: :::
932 ..... IloThValSerAspArgIleu 939
824 GAATGTGGTGCACCACAATGTTCCAGAACCAAGAATGGGAGCGTTAA 775
      || :||: ::: |||:::||::: |
940 GILeuAlathrValGIuarGpheAnaIathAlaLeuGLyGLy 956
774 AACAGTGTTGCCGCGACGAACAAACCTGCTGAGCTGGCACCATCGTTAC 725
      | |||||
956 sleuglYLyIeuthrPheaspGlySeurSerIleuProLysI 973
724 TCAGAAAAGATGTCAGAACCGGAGAAATATCATTCAT.....GTRACA 681
      |||||::: ::: |||||
973 lEgLYLSArSserAlValGlunSPseuLeuPheasnLYSLValVthr 989
680 ACTGATATCAACACCCCTTGAATTAGCGATGATATATCAACTTGGCCAGT 631
990 SerGIyleuGLIThrValASpASPspyrILYSscYSerSerGLYT 1006
630 TCAATTACGTCACCGTGACCTTCGACCATATNATGACCAACCTGTAATAC 581
      |||||::: ::: |||||::: ::: |||||
1006 rASpVALAAspleuValcYsalagInTYTLasmoLYlleMeValI 1023
580 TTGCGTAGCATACMAATGGATTCGATGTCCATGCATTGGCTGTCATG 531
      || :||| ::||| ||||| ::: |||||
1023 euProclYalValAspGLYasnLYsmetSetMetLYThrAlaserIeu 1039
530 TTATNGGTTTTAAGCAT...GCATTGATTTGGTGGCGCATATTATACAT 484
      ::: |||||::: ::::: |||||::: |||
1040 lEGlyGLymeALaleuGLyserIethrSerAlavalAlaval..PROP 1056beuLn
483 TT.....CGTTAAATTTGCTCCAATCAGA 458
1056 heAlamEtGINvalGlnAlarGLueAnstLY-ValAlaleuGln 1070
```

```

SQ      Sequence      374 AA:

alignment_scores:
    quality:      89.50      length:      363
    ratio:        0.533      gaps:        18
    percent similarity: 46.281      percent identity: 21.763

alignment_block:
US-09-323-427-2/rev x W88310      ..

Align seg 1/1      to: W88310      from: 1      to: 374

1552      TTGCACATGCTGTGCAGAGGTGAGCCAGAAATTGATGTGGACCAACTTCA      1503
|||||      ::::::::::::::::::::      |||:::
38      LeuThrGlyIleSerGluValIlyProThrGlnAsnIleAsn.....      51
1502      ATACACATCAATTTTAATACACGTATGATTCGATTCGAGGAC.....A      1462
|||||      :::::
52      .....IleHisTyrValAsnMetAspLysAsnPheArg      63
1461      TGTTAATGCGAAGGCTTATGATTCACAGAGGTGCCGTAAATGATGAC      1412
|||||      ::::::::::::::::::::      |||:::
63      ePhePheArgAlaLeuPheGlnValIlySerIleIleValAlaLeuLys      79
1411      GTCGACGTCAAGTTCGCCGAATTTCACTTCATTTGATTCATGCAAGTT      1362
|||||      ||:::|||||
80      ProAspIle.....IleHisSerHisMetPheHisAlaAsn..      91
1361      GCGCGTACACGATCTCGATTCACACGTGATTTTGTACACAGAACHTGT      1312
92      .....I      92
1311      TGTGATTTGGTTTCACTCATTTATTTGTACCAAGTTG...ATTCGTCAT      1265
|||||      ||::::::::::::      ||:::
92      LePheSerAlaPheIleArgMetLeuIleProAlaValProLeuIleCys      108
1264      ATGAGTACACATGCTTTTACATGCAAGCTGATTAACAGGTTGTGCACAG      1215
|||||      :::::
109      ThrAlaHisAsnLysAsnGluGlyAsnAlaArgMetPheCysTyrTrp      125
1214      ATTGAGGTATCTGAAATTCACACAGCTTTTCAACTGCACAAATTTGCCGAT      1165
|||||      ||::::::::::::      ||:::
125      Gln.....SerAspPheLeuAlaSerIleThrThrAsnValS      138
1164      GCCGATATGCCGTATGAAATTTTGGATGCTGACACACCGGTCACACAG      1115
|||||      ||::::::::::::      ||:::
138      eLysGluAlaValGlnGluPhe.....IleAlaArgLysAlaThrPro      152
1114      TTCAATTTGTCATCATGTGCTCAGCCAGTTTATCAT.....      1080
|||||      :::::
153      LysAsnLysIleValGluIleProAsnPheIleAsnThrAsnLysPheAs      169
1079      .....AAATGACATGCGATTTCTGAAACCGTTGATA      1049
|||||      |||      |||      ||:::
169      pPheAspIleAsnValArgLysLysThrArgAspAlaPheAsnLysAs      186
1048      CTTTGTGCCGCTTGTCCATTCCTGCTTGTGTCATGATGTAACGGTCAT      999
|||||      :::::
186      sPseThrAlaValAlleuLeuAlaVal.....GlyArg      196
998      ACTGTGGAA.....ATTCCTAAATGCTGATGATGTC      967
|||||      ||:::
197      LeuValGluAlaLysAspTyrProAsnIleuAsn.....Al      209
966      TCTTGATTAATATTGCTAAATATATTGGATATTCACACAGAT.....      924
|||||      ::::::::::::::      ||:::
209      aIleAsnHisLeuIleLeuSerLysThrSerAsnCysAsnAspPheIle      226
923      ..TTAATGGCTGGCCACAGAGCTACAGTATACAAATATGCGGATGATCA      876
|||||      |||      |||
226      eLeuIleAlaGlyAspLysAlaLeuArgAsnLysLeuLeuAsp.....      240

```

```

875 CAGCTTTTCTATCATGCGCAGATCAGTATTACATTAA..... 837
      |||||:.....: ||
241 .....LeuValCysGlnLeuAsnLeuValAspLysValIlePhePhe 254
      :|||:.....:
836 ..GAACCAATAGCGAATGTGTTCACCAATGT..... 804
      :|||:.....:
254 uGlyGlnArgSerAspIleuysGluLeuMetCysAlaIleAspLeuPheV 271
      |||||:.....:GCTGTTAAACAGGTGGT 765
803 .....TCAGAACCAAGGATTGGA...GCTGTTAAACAGGTGGT 765
      |||||:.....:
271 aLeuSerSerGluTTPGluGlyPheGlyLeuValAlaGluAlaMet 287
      :|||:.....:
764 GCCGACCAAAACCTGCTGCAGCTGCACCACTTCGTTACTCAAGAAAG 715
      |||:.....:
288 AlacysGlnArgProValAlaAlaThrAspSerGlyGlyValLysGluVa 304
      :|||:.....:
714 ATTCGCAAGAACCGAGATATCATGTATGATGCAATGATATC..... 672
      :|||:.....:
304 lValGlyProHisAsnAspValIleProValSerAsnHisIleLeuLeuA 321
      :|||:.....:
671 .....AACACCTTGAAATTAGCGATGATATCAAGCTTGG 636
      :|||:.....:
321 lGluGlyIleAlaGluThrLeuLysIle...AspAsnAlaAlaArgLys 336
      :|||:.....:
635 CCAGTTGATTTACGTACCGCTGCACCTTCGTGCACATTAAT 597
      :|||:.....:
337 lIleIleGlyMetLysAsnArgGluTyrIleValSerAsn 349
      :|||:.....:

```

seq_name: A_Geneseq_36:R59926

```

seq_documentation_block:
ID   R59926 standard; protein; 3079 AA.
AC   R59926:
DE   22-FEB-1995 (first entry)
DE   GAP protein Iraz.
KW   Ras; GTPase activating protein; GAP; GAP related domain; GRD;
KW   RAS2; v-Ras; heat shock; neurofibromatosis type 1; NF1.
OS   Saccharomyces cerevisiae.
PN   M09416069-A.
PD   21-JUL-1994.
PE   12-JAN-1994; 000198.
PR   15-JAN-1993; US-004824.
PA   (SCHE ) SCHERING CORP.
PI   Kaziro Y, Nakafuku M;
PI   WPI: 94-249216/30.
PT   Blocking Ras-induced effects on a cell - by introducing a GTPase
PT   activating protein to the cell, used esp. in treatment of cancers
PS   Disclosure: Page 63-72; 87pp; English.
CC   Human neurofibromatosis type 1 (NF1)-GAP related domain (GRD)
CC   mutant clones NF201 (given in R59221) and NF204 (R59922) show
CC   v-Ras-induced transformation activity for RAS2Val19, and inhibit
CC   strong suppression activity in mammalian cells. The mutation
CC   sites of these proteins were located in one of the most conserved
CC   regions of GRD. These sites were compared with those of other
CC   GRD family proteins, yeast Iraz2 (R59926) and Iraz1 (R59923),
CC   human GAP (R59924) and Schizosaccharomyces pombe Gap1 (R59925).
SQ   Sequence 3079 AA:

```

```

alignment_scores:
  Quality: 89.00      Length: 752
  Ratio: 0.264      Gaps: 42
  Percent Similarity: 44.814      Percent Identity: 18.617

```

alignment_block:
US-09-323-427-2/rev x R59926 ..

Align seg 1/1 to: R59926 from: 1 to: 3079

```

1742 ACTGACAGACAAACAACAAC.....AACACAAACAACAACAAC 1699
      |||:.....: |||||
2300 ThrSerSerProAsnSerAsnLysValLysGlnGlnLysGluArgTh 2316

```

```

1698 AGCAATAATATACCCCATCATAGTGGAGAGACAGACAGAACAAATCTTAG 1649
      |.....:
      :|||:
2316 rIleLeuLeuCys.....HisLeuLeuV 2325
      :|||:
1648 TTTTCTTAAATAATGAAATTTCATTAATCTTCTGAATG.....ATG 1608
      |||||:.....:
2325 aISerLeuIleGlyLeuPheAspIleSerLysMetLysAsnSerSer 2341
      :|||:.....:
1607 ATTGCTTATTTGCTTTCTGTACTACACTTAATGCAATGTCTTATTCG.. 1560
      :|||:.....:
2342 TyrAsnLeuIleAlaIleThrGlnAlaSerPheGlyLeuAsnIleGlySe 2358
      :|||:.....:
1559 .....ATTCGGTTGACAAT..... 1545
      :|||:.....:
2358 rHisPheHisArgSerProGluValTyrValProGluAspThrThrThrP 2375
      :|||:.....:
1544 .....GGTGTGAGAGTGAGCGACAGAAATTGAATGTGGACCACTTCATA 1500
      |||||:.....:
2375 heLeuGlyValIleGlyLysSerLeuAlaGluSerAsnProGluLeuThr 2391
      :|||:.....:
1499 ACAATCAATTTTAAATACAGCAATGATGATTCG...AAGACATGTTTATGT 1453
      :|||:.....:
2392 AlatyMetPhe.IleTyrValLeuGluAlaLeuLysAsnAsnValIleP 2408
      :|||:.....:
1452 GAAAGCTCTTATGATGATCAGACAGGTGGC..... 1424
      :|||:.....:
2408 rHisValTyrIleProHisThrIleCysGlyLeuSerTyrTrpIlePro 2424
      :|||:.....:
1423 .....GTAATGATGAGGTGAGGTGACGTCAAGTTGCCGGAATT 1389
      :|||:.....:
2425 AsnLeuTyrGlnHisValTyrLeuAlaAspAspGluGluLysProGluAs 2441
      :|||:.....:
1388 TCACCTTCATTTGATTCATCAATGATGTCGC.....GTACAGATCTCT 1345
      :|||:.....:
2441 nIleSerHisIlePheArgIleLeuIleArgLeuSerValArgGluThra 2458
      :|||:.....:
1344 GAATCCACGCTGATTTTGTAAACAACACTGTTCATTT..... 1304
      :|||:.....:
2458 spHeLysAlaValTyrMetGlnTyrValTrpLeuLeuLeuAspAsp 2474
      :|||:.....:
1303 .....CGTTTCATCCATATTTGTACCAAGTTGATGCTGAT...ATCGA 1260
      :|||:.....:
2475 GlyArgLeuThrAspIleIleValAspGluValIleAsnHisAlaLeuGl 2491
      :|||:.....:
1259 GTACATGCTTTTACA.....TGAGAG..... 1238
      :|||:.....:
2491 uArgAspSerGluAsnArgAspTrpLysThrIleSerLeuLeuThrV 2508
      :|||:.....:
1237 .....CTGATTAACAGTTAGTCAC 1217
      :|||:.....:
2508 aLeuProThrThrGluValAlaAsnAsnIleIleGlnLysIleLeuAla 2524
      :|||:.....:
1216 AGATTGAGGATGTGAATACACACATGCTTTTCAACTCAATATGTCGG 1167
      :|||:.....:
2525 LysIleArg.....SerPheLeuProSerLeuLysLeuLysLysLysLys 2536
      :|||:.....:
1166 ATGCCAGTATGCCGTTATGAATTTTGATGTGTGGACCAACGCGTCAAC 1117
      :|||:.....:
2537 .....AlaMetThrGlnSerTrp.....SerG 2544
      :|||:.....:
1116 ACTTCATTTGCTATCATTTGCTGACGCACTTATATCAATAATGACATGCG 1067
      :|||:.....:
2544 lLeuThrIleLeuValLysIleSerIleHisValIlePheGluThrSer 2560
      :|||:.....:
1066 ATTTCGAACCGGTGATCTTCTGTGCGGTTGTGCATTCCTGCTTTCG 1017
      :|||:.....:
2561 LeuLeuValGlnMetTyrLeuProGluIleLeuPheIle.....ValSe 2575
      :|||:.....:
1016 GATGATGTAGCGTGATACCTGTGGAATTTAAATGCTGATGATGTGCG 967
      :|||:.....:
2575 rLeuLeuIleAspValGlyProArgGluLeuLysArgSerLeuHisGlnL 2592
      :|||:.....:
966 TCTTGATAATATTTTGCTTAAATAATTTGGAATATCCACAGATTAATGG 917

```

```

|||||:|||||: 2597
2592 euLeuMetAsnValGys..... 2597
916 CTGGCCAGAAAGCTCAGTATACAAATATGCCGATCGATCAGAGCTTTTC 867
2598 .....HisSerLeuNI 2601
866 TATCAATGCCAGATCAGTATACCAATTAAG.....AAC 832
2601 aileAsn.....SerAlaLeuProGlnAspHisArgAsnAsnLeuAspG 2616
831 AATATCAGATGTGTGCACCAATGTTCCAGAC..... 797
2616 IuileSerAspIlePheAlaHisGlnVallySphMetPheGlyPhe 2632
796 CACAGAGATTCGAGAGCTGTTAAACAGGTGTCGCGAG..... 758
2633 SerGlnAspYsglyArgIleLeuGlnIlePheSerAlaSerSerPheAl 2649
757 .....CAAACTGCTGCAGCTGC 739
2649 aserlySphAsnIleLeuAspPhePheIleAsnAsnIleLeuLeuNI 2666
738 GCAGCTTC.....GTTTACTCAAGAAAAGATCTGCAGACCGAGAGATA 695
2666 etGluTYrSerSerThrTYrGlnAlaAsnValTYrPlyThrArgTYrLys 2682
694 TCATTTGATGTACAGATGATATCAACA.....CCCTTGAA 660
2683 LysTYrValLeuGlnSerValPheThrSerAsnSerPheLeuSerAlaAr 2699
659 ATTACCGATGATTAATCAAGCTTTCGACATGTTACGTCACCGTGCAC 610
2699 gserIleMetIleValIleGlyIleMetGlyLysSerTYrIleThrGlnGlyL 2716
609 TCTGCACATATG.....GACAACTGTTAACTGTCGACAGTACAAA 566
2716 euCysLysAlaMetLeuIleGlnThrMetLysValIleAlaGlnProLys 2732
565 ATGCAATCT.....GCATGTCCACAT..... 545
2733 IleThrAspGlnHisLeuPheLeuAlaIleSerHisIlePheThrLysE 2749
544 .....TTGGCTTCTCAATGTT 529
2749 rLysIleValGlnGlyLeuAspProAsnLeuAspLeuMetLysHisLeuP 2766
528 TATGGGTTTAAAGCATTG.....CATGATGCTGCAGCTGCA 494
2766 hetrPheSerThrLeuPheLeuGlnSerArgHis.....ProIle 2779
493 TTATTACCAATTCGTTAAATTTCCGTCGAA.....ATCAGAAAG..... 455
2780 IlePheGlnGlyAlaLeuLeuPheValSerAsnCysIleArgArgLeuTY 2796
454 .....CATAAATATATGTTAGATCATCAGACCAATAA 421
2796 rMetAlaGlnPheGlnAsnGlnSerGlnThrSerLeuIleSerThrLeuL 2813
420 TA.....AACTGCCATATATATTCGTTTCTTATCAATCCTTCT 380
2813 euLysGlyArgLysPheAlaHisThrPheLeuSerLysIleGlnAsn... 2828
379 AATACTAATTTTACCTACAAATATATATGATGTAGCAAAATATATCTG 330
2829 .....leuSerGlyIleValIlePrasngLysAspAsnPhen 2840
329 TAATACA.....ATAAGTATATTTTCATCA 304
2840 rHisIleLeuIlePheIleIleAsnLysGlyLeuSerAsnPropheIleL 2857
303 AACTTCTCTATCGCTTTATA..... 281

```

```

2857 ySerThrAlaPheAspPheLeuLysMetMetPheArgAsnSerTYrPhe 2873
280 .....GCTTCTGAAAAGCTTATTCATTAT...TCAGTAATCTTTTA 243
2874 GlnHisGlnIleAsnGlnLysSerAspHisTYrLeuCysTYrMetPhe 2890
242 TATGATACTATTTGTAATGTTTCATCA.....TTAGCCCATG 205
2890 uLeuTYrPheValIleLeuAsnCysAsnGlnPheGlnGluLeuLeuGlyAspV 2907
204 AATAGTTCGCT.....TTGTTATTCATTCATTATCAACTGCTCT 164
2907 AlaSphGlnGlyGlnMetValAsnIleGlnLysAsnThrIlePro 2923
163 ATTTATTTTACAGATTTATTCATTGTCATAT...ATCAGAAATTTATAC 117
2924 LysIleLeuLeuGlnLutPheLeuSerSerAspAsnGlnAsnAlaHisnIleH 2940
116 CTGTGAT.....TGCCCATTT..... 101
2940 rLeuTYrGlnGlyAlaIleLeuPheLysCysSerValThrAspGlnProS 2957
100 .....TTATGCGCATTCATTCTATTCGTAACATTCACCTT 62
2957 etArgPheArgPheAlaLeuIleIleArgHisLeuLeuThrLysLysPro 2973
61 ATTTCATTTATTCGAATTAATAAGTATTTCTTTGTGAAAAAATAAAAA 12
2974 IleCys.....AlaLeuArgPheTYrSerValIleArgAsnGlnIleAr 2988
11 AAAA 8
2988 gLys 2989

```

```

seq_name: A_Geneseq_36:R42452
seq_documentation_block:
ID R42452 standard; protein: 1634 AA.
AC R42452:
DT 27-MAY-1994 (first entry)
DE Enzyme involved in eicosapentaenoic acid (EPA) synthesis.
KW EPA: eicosapentaenoic acid synthetase; drug; anticoagulant;
KW hypolipemic; hypoglycemic; antihypertensive; anticancer; pesticide;
KW foodstuff; additive.
OS Shewanella putrefaciens.
PN WO9323545-A.
PD 25-NOV-1993.
PR 14-MAY-1993: JP-147945.
PR 15-MAY-1992: JP-147945.
PA (SAGA ) SAGAMI CHEM RES CENTRE.
PI Kato S, Kondo K, Yamada A, Yazawa K;
DR WPI: 93-386577/48.
DR N-PSDB: 051128.
PT Gene coding for eicosapentaenoic acid synthetase - is isolated
PT from pseudomonas, Alteromonas or Shewanella and used for
PT recombinant prodn. of eicosapentaenoic acid
PS Claim 6: Page 52-63; 106pp; Japanese.
CC EPA is useful as a drug, having anticoagulant, hypolipemic,
CC hypoglycemic, antihypertensive and anticancer activity. It is also
CC a pesticide and is useful as a nutritional foodstuff and animal feed
CC additive.
CC Sequence 1634 AA:
SQ

```

```

alignment_scores:
  Quality: 88.00      Length: 500
  Ratio: 0.376       Gaps: 23
  Percent Similarity: 46.800  Percent Identity: 18.200

```

alignment_block:

US-09-323-427-2/rev x R42452

Align seg 1/1 to: R42452 from: 1 to: 1634

```

1745 CCGACTGCAGCACAACAACAACAACAACAACAACAACAACAAGC 1696
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
827 ProPheAlaIalValIleAspAlaIalValSerPheThrIleSerArgAl 843
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1695 AATAATAACCCATCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1646
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
843 AleuTyrSerAsnAlaThrIleGlyGly...LeuTyrGluSerThrAlaIal 859
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1645 TTCATAAATGCAATTTACTAAATCTTCTGAATGATCATTCGTTATT 1596
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
859 ySileuValSerPheIlySthS...MetLeuGlnSerVal 872
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1595 GCTTCTGCTACACTTATTCGATTGCTTATTCGATTCCGGTTGACA 1546
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
873 ArgPheThrSerGlnLeuGlnAlaMet...TyrAsnAs 884
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1545 TGGTGTGAGAGTGACCCAGAAATGTAATGTGACCAATTCATAA... 1500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
884 pGly...AlaArgValPheValGlnPheGlyProIlyAsnIleLeuG 899
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1499 .....ACA 1497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
899 IuIlySleuValGlnGlyThrIleuValAsnThrGluAsnGluValCysThr 915
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1496 ATCAATTTTATACACGTAATGCAATTCAGAGACATGTTATGTGAAGG 1447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
916 IleSerIleAsnProAsnProIlyValAspSerIleuGlnLeuS... 931
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1446 TCTTATGATCACAAGGTTGCCGTAATGATGAAGTGAGCGTCAAGTTG 1397
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
932 .....GlnAlaIleMetGlnLeuA 938
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1396 CC.....GGAATTTCACTT.....CCATTGATTGATGCAAT 1365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
938 IalValThrGlyValIleValLeuSerGlnIleAspProIlyGlnAla...Asp 953
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1364 GTTGGCGGTACACGATCTGGAATCCAGTGTGATT..... 1329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
954 IleAlaIalProAlaIlySlySerPrometSerIleSerLeuAsnAlaI 970
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1329 ..... 1329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
970 aAsnHisIleSerIlySAlaThrArgAlaIlySmetAlaIlySerLeuGln 987
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1328 .....TTTGAACAACAACGTGTT.....GTCAATTCGTTTCACTCA 1290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
987 hrcIlyIleValIthrSerGlnIleGlnHisValIleGlnGluIlySileVal 1003
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1289 TTTGTTACCAAGTGTGATCGATATCGAGTACAATGCTTTTACATGCA 1240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1004 GluValGlnIlySleuValGlnValIleValIleValIleValIleVal 1020
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1239 AGGTGATAAAGAGTATGACACAGATGAGATATCTGAATC...ACAA 1193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1020 uValGlnIlySleuValGlnValIleValIleValIleValIleValIle 1037
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1192 GTGCTTTCAAACTCAATGTCCCGATGACAGTATGCGTTATGAAT 1143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1037 snAlaIleGlnThrArgSerValIleAlaIleProValIleGlnAsnGlnVal 1053
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1142 TTGGATGTGAGACCAACCGGTCAACAGTTCATTTGCTATCATTTGTC 1093
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1054 Val.....SerIlySAsnSerIlySProAlaValIleGlnSerIleSer 1068
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1092 GCCAGTTTATCATTAATGAGATGCGATTCGAACCGTTGATACTTCT 1043
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1068 palAlaIleuSerAsnPheAlaIleGlnIleGlnIleThr..... 1080
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1042 GCGCGGTGTTCATTCCTGCTTT.....GTGAGATGATGTAACGGTAT 999
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1081 ..AlaGlnIleuHisGlnIlePheLeuAlaIleProGlnIleTyrGln 1096
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

998 ACTGTGAATTTCTAAATGCTGAT.....GGATG 970
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1097 ThrPheThrThrLeuMetThrGlnIleAlaIlySleuAlaIleSerSerGlyVa 1113
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
969 TGCCTGTGATAAATTTGTGCTAAATATTTGGA..... 936
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1113 lAlaIleProGlnSerLeuGlnArgSerMetGlnIlePheHisGlnLeuG 1130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
935 .....TATCCAAACAGATTTAATGCGCGC 912
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1130 lAlaGlnThrLeuGlnIleSerHisThrGlnPheLeuGlnMetGlnAlaIle 1146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
911 CAAGAGCTCGCTATACAAATATGCGGATGATCAGACGTTTCTATCA 862
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1147 SerAsnIleAlaIleValAsnLeuAsnLeuAsnSerSerGlnAlaIleThr 1163
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
861 ATGCCAGATCAGTATTTACATTAAAGAACCAATATGCGATGTGTGAC 812
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1163 aProAlaIle.....HisAsnGlnAlaIleGlns 1173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
811 CACAATGTTCAAGACCAAGGATTCGAGCTGTTAAACAGTGTGCGC 762
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1173 ergIleValIleGlnSerGln.....Thr 1180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
761 GCACAAACCTGCTGCACTGCGCAACTTCGTTTACTCAAGAAAGA... 714
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1181 AlValGlnProValIleSerThrGlnValAsnHisValSerGlnIlePro 1197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
713 .....TCTGCAGAACCGGAGAAATATCATTCATGATGATAC 683
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1197 oThrGlnAlaProAlaProIlySAlaGlnPro.....AlaProVal 1211
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
682 GAACGTATATCAACACCCCTGAAATTAGCGATGATATCAAGCTTTGCA 633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1211 hrThrAlaValGlnThrAlaProAlaGlnValValArgGlnAlaIlePro 1227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
632 GTTGATTTAGTCAACCGTGCACCTTCGCAACATATGAGACAACCTGTA 583
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1228 Val.....GlnAlaAlaIleGlnProIleAs 1236
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
582 ACTTGTCAGTACAAANAGATGTCATGTCACCATTTGGCTTTCGA 533
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1236 nThrSerValAlaIleThr.....ThrProSerAlaPheSerA 1249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
532 TGTTATGGGTTTAACATTTGATGATGCTGCGCTCATTTATTACCATT 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1249 IagIuThrAlaIleuSerAlaThrIlyValGlnAlaIleThrMetLeuGlnVal 1265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Align seg 1/1 to: US-08-415-751-6 from: 1 to: 362

```

990 TTCACAGTATCACCGTTACATCATGCAAGAGAGATGAC..... 1034
22 PheHisSerLeuSerHisLeuIleuAspIlySlnPheThrSerPheTr 38
1035 .....AACGGCGAGAAAGTATCAACGGTTTCAGATCGCATGTC 1074
38 pIleGluArgAsnGlnThrAsp..... 45
1075 CATTTATGTAATACTGCTGACCAATGATAGCAATGTAACGTGTTGAC 1124
46 .....TrrIleGlnTrpValThrSerGluCys...IleAsnTrp***Thr 59
1125 GGTGGTGCACATCCAA.....AATTTCATACCGCATACGCGCATCG 1168
60 ValValSerSerValSerThrValIysPheLeuAlaIleArgIlySlnPth 76
1169 GGACATTTGAGTTTGAAGAAGCATTTGATTCAGATACCTCATCTGT 1218
76 rAspIlySProAspIle..... 81
1219 GCACATACTGTTTATCAGCTTCATGTAAGCAATGTTACTGATATGC 1268
82 .....Val***LeuValIlyGly...Cys 88
1269 AGC.....ATCACTTTGTTAC.....AAATATGAGAGAAAGCAA 1306
89 ThrIlySleuProAspPheGlySerAsnGlyArgProThrIleGlySerI 1305
1307 TGACACAGAGTTGTTGTACAAAATACCACGCTGATTCAGAGATCGTGA 1356
105 eGlyIysProValIleuLeu***Val***ThrTrpGlu***IleAlaCysI 1322
1357 CGCGCAACATTCGATGAATCAAA...TGAAGTGAATTCGCGCAACTGG 1403
122 leSerTyrLeuSerCysLeuArgTrp..... 131
1404 AGCTCCACCTTCATCAT.....ACGGCAACCTTCGATCATATAA 1444
132 ...AsnTrpPheIleIleArgVal*****ThrThValIleVal... 146
1445 GACCTTTCATTAACATGCTTCGATGATGATTCAGTGTATTAATG 1494
147 .....IleAsnHisSer.....CysValIleuIle 155
1495 ATTGTTATTGAAGTGGTCCACATTCATTTCTGGCTCACCTCGACACC 1544
155 IlnHisLeuPheAsnTrpAspTrpCysAsnPhe..... 165
1545 ATTGTCAACCGGAATCGAATAGCATGCAATAGTAGTACAGAAAG 1594
166 .....AsnValThrAspIleArg***TrpCysArgCysCysCys... 178
1595 CAAATAGACGATCATCTTCAGAGATTTAGTAATTCGATTTTGA 1644
179 .....HisPhe..... 180
1645 AAAACTAGATTGCTCTGCTGCTCTTCCTCCACTGATGGGTTATAT 1694
181 .....ValSerLeuGlySerSerSerCysArgTrp...HisCys 192
1695 TGCCTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1733
193 CysCysCysCysCysCysCysCysCysCysCysCysCysCysAsnTrp 209
1734 ....TGCAGCAGTCGGTGT 1748
209 u**CysCysCysCysCys 215

```

seq_name: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:US-08-415-751-35
seq_documentation_block:

```

; Sequence 35, Application US/08415751
; Patent No. 5643772
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: POLYPEPTIDES BINDING ANTI-
; TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA
; TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID
; TITLE OF INVENTION: VECTOR AND TRANSFORMED HOST AND
; TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND
; TITLE OF INVENTION: DIAGNOSIS AND KIT
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESSES:
; ADDRESS: PHILLIPS, MOORE, LEMPJO & FINLEY
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: California
; COUNTRY: United States of America
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette - 3.5 inch, 1.44 KB storage
; COMPUTER: PC
; OPERATING SYSTEM: DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/415,751
; FILING DATE: 03-APR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/071,880
; FILING DATE: June 1, 1993
; APPLICATION NUMBER: 07/891,301
; FILING DATE: May 29, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hana Dolezalova
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.19-2 (HND)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-1677
; TELEFAX: (415) 324-1678
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 362 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORGANISM: Cryptosporidium parvum
; FEATURE:
; NAME/KEY: Positions coded by nonsense codons are
; NAME/KEY: identified as xaa.
; US-08-415-751-35

```

alignment_scores: Length: 31
Quality: 109.00
Ratio: 6.056
Percent Similarity: 58.065 Percent Identity: 51.613

alignment_block:

US-09-323-427-2 x US-08-415-751-35 ..

Align seg 1/1 to: US-08-415-751-35 from: 1 to: 362

```

1656 TGCCTCCTGCTCTTCTCCACATGATGAGGATTAATTCGCTGTTGTTG 1705
135 CysCysCysCysLeuAsnTrpThrLeuTrpTrpIlyrGlyCysCysCys 151
1706 TTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1748

```

```

151 scys***LeuTrpIyAsnGysCysCysCysCysphecysCys 165
seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-477-451-8
seq_documentation_block:
; Sequence 8, Application US/08477451
; Patent No. 5928865
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori CagI Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,451
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0335,002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2708
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3200 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-477-451-8

alignment_scores:
Quality: 101.00 Length: 453
Ratio: 0.537 Gaps: 24
Percent Similarity: 41.501 Percent Identity: 22.075

alignment_block:
US-09-323-427-2 x US-08-477-451-8 ..
Align seg 1/1 to: US-08-477-451-8 from: 1 to: 3200

564 ATTTGACTGACGAAGTAT.....TACAGTTGCCATTTATGTT 604
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
745 ILeValphecysArgIlePheGluLeuValTyrArgLeuLeuLeuLeu 761
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
605 GCAGAAAGTCGACGCGTAAATCAACTGGCAAGCTGATTATATCATC 654
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
761 slYsTrhCys.....PheSerPheTyrArgLeuIleGluIleVal 774
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
655 CTAAATTCGAAGSGTGTGATATC.....AGTTCGACATCAATGATATT 698
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
774 AlAsnPheLeuSerGlnPheLeuIleProTyrLeuTrhGluIle 790
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
699 CTCGCGTTCTGCAGATCTTTCTTGAGTAAACGAGTTGGCAGCGCAG 748
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
791 IleAsnPheIleMetProPhe.....GlnArgGluAlaArgCysLe 804
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
749 CAGGTTTTCGCGGACACACCTGTTTAAACAGCTCCGAATCCTTGTTG 798
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
804 uProTyrAsnTrhGlnTrhCysTyr..... 813

```

```

799 TCGAACAATTGGTTCGAACAACTTCGCTATTGGTCTTTAATGTAAT 848
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
814 .....LeuIleGluSerValAsnPheCysIleTyrLeuPheSer...Asn 827
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
849 ACTGATCTGGCGATGATAGAAAGACTGTGATCATCCGCAATTGTTATA 898
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
828 LeuSerValAsnGlyPhePheLeuYsAlaSerCysLeuProCysPheLeuPh 844
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
899 CGTGAGCTCTTGGCCAGCCATTAATCTGTGATATTCGAATTTATT 948
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
844 eIleAsnPhe..... 847
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
949 AGCAATATTTATTCAGACATCCATCCATTTAGAAATTCACAGT 998
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
848 .....SerIleAsnGlyPheValPheGluSer 856
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
999 ATCAGCTTACATCATCGAC.....AAAGCAGCAAT 1030
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
857 IleIleValIleValIleGluIleSerIysPheCysIleGlnLeuPheLe 873
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1031 GGACAAACCGCGCAGAAAGTATCAACGGTTTCAGAAATCGCATTCATTTA 1080
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
873 uAspValArg.....LysValPheArgLysSerCysGlyPheC 886
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1081 TGATTAACCTGGCTGACCAATGATAGCAAAATTTGAAGTGGTACCGGT... 1127
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
886 YsLeuArgProIleTrhAspLysSerIleHisTrhIleLeuAsnGlyLeu 902
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1128 .....TGCTCACCATCCAAATTTCAATTAACGGCATCTGTG 1162
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
903 IleTyrPheArgPhePheSerSerIleGluTrhPheIlePhe..... 916
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1163 GCATCGGAGCAATTTGAGTTTGAAGACAGTGT..... 1196
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
917 .....LeuSerLeuLeuGlyGlnLeuLeuGluIlePheG 929
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1197 .....GATTTCAGATACCTCAATCTGTGCATACTGTTTATTCAGC 1238
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
929 LuIleGlyArgLeuLeuTyrProTrhGluIleTrhArgCysPheSerHis 945
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1239 TTCCATTAATAAGCATTTGTACTGATATGCACATCACTTTGGTAAACA 1288
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
946 TyrArgLeu.....IleGluIleIleAsnPheIleArg... 956
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1289 ATAATGATGAACGAATGACAAATGACAACTTTGTTACAAAATACCACT 1338
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
957 .....TyrLeuLeuIleAspAspHis.....IlePheAsnGlyIleP 969
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1339 GGATTCAGAGATCGTGTACCGCGCAACATTCATGATCAATGGAAGTGA 1388
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
969 heLysValAlaPheCysLeuArgTyrLeuAlaGluIle..... 981
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1389 AATTCCGGCAACTTGGACGTCCACCTTCATCATTAACGCAACCTTCTGAT 1438
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
982 .....LeuIlePheAlaIlePheGluPh 989
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1439 CATAAGACCTTCACATAAACAATGTCCTTCGAATGATCATAGCTGATTA 1488
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
989 eIleGln.....SerCysIleL 995
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
995 YsVal.....PheAsnPheAsnProPheTrpIle... 1004
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1536 TTGACACCATTTGTCAACCGGAATCGAATTAAGCAATGCAATAGCTGAG 1585
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1005 ...AspTrhIlePheHisArgTrhPheLysGlyPheCysPhePheCysLe 1020
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1586 TACGAAAGCAATTAAGACGAATCATTCATTCAGAGATTTAGTAATTTG 1635
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1020 uAlaGlu.....IleLeuIlePheAlaIlePheG 1030
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-07-707-367-2

```

US-07-707-367-2

alignment_scores:
    Quality: 100.00
    Ratio: 0.541
    Percent Similarity: 47.315
    Percent Identity: 19.437

alignment_block:
    US-09-323-427-2/rev x US-07-707-367-2
    Align seg 1/1 to: US-07-707-367-2 from: 1 to: 935

1517 TGTGGACCACTTCATATACATCAATTTATTAACCGTAATTCATTCGA 1468
    ||| ||||| ||||| ||| ||| |||
526 CysGlnProHisValAlaValAspProIleThiGlyAsnIlePheVa 542
AGCACAATTATTAATGGAAGAAGCTTATGATCAAGAAGTTGCCGTAATG 1418
    :||
542 LAlAspGlyTyr..... 546
1417 ATGAAGGTGGACGTCAAGTTGCCGAATTCACCTTCATTTGATTCATG 1368
    |||
547 .....Cys 547
1367 AATGTTGCCGTCACAGCATCTGCATCCAGCTGTAATTTTGA..... 1323
    ||| ||||| .....||| ||||| |||||
548 Asn...SerArgIleMetGlnPheSerProHisnIleMetIleMetG 563
1323 ..... 1323
563 nTTPGlyGlnGluThiSerSerAlaProArgProGlyGlnPheArgI 580
1322 .....ACAACTGTTGTCACTTCGTTTCATCCATTAATTTATTCAC 1281
    :: ||||| ||| |||||
580 IeProHisSerIleThiMetValProAspGlnGlyGlnLeucCysValAla 596
1280 AAGTTGATGTCGATGTCATGTCAGTACATGCTTTTACATGCAAGCTGATA 1231
    :||| ||||| ||||| ||||| |||||
597 AspArgGlnAsnGly...ArgIleGlnCysPheHisAlaGluThiGlyAs 612
1230 AACAGTTAGTGCACAGATTGAGGATTCGAATC...ACAACTGCTTTG 1184
    ||||| ||||| ||||| |||||
612 nPheValLys...GlnIleLysHisGlnIleGluGlyArgGlnValPhe 628
1183 AAATCAATGTGCCGATGCCAGTATGCCGTTATGAATTTTGGATGCT 1134
    :|||
628 IAlaIleTyrAlaPro.....Gly 634
1133 GGACCAACCGGTCACACAGTTCAATTTGCTATTCATTTGGTCACAGTTTA 1084
    ||| ||||| ||||| ||||| |||||
635 Gly.....ValLeuTyrAlaValAlaHisnIleLysProTyrTyr 646
1083 TCAATAATGACATGCGATTTGGAACCGTTGACTTTCTGCGCGGTTG 1034
    |
646 I.....GlyTyrSerAlaProV 652
1033 TGCATTCCTGCTTTGCGATGATGATGAACGGTACTGTCGGAATTTCTA 984
    ||||| ||||| ||||| ||||| |||||
652 AlGlnIlePheMetLeuAsnIlePheSerHisnGlyAsp..... 663
983 AATGCGATGATGATGCTGTTGATAAATATTG.....CTAAATAAATT 940
    ||||| ||||| ||||| |||||
664 .....IleLeuAspThrPheIleProAlaArgLysAsnIle 675
939 GGAATATCCAAACAGATTAAATGCTGGCCAGAAAGCTCACTATCAAT 890
    ||||| ||||| ||||| ||||| |||||
675 eAspMetProHisAspIleAlaAlaAlaAspGlyHisValGlyValG 692
889 ATGCGGATCCATGACAGCTTTTATCAATGCGCAATCACT..... 849
    ||||| ||||| ||||| |||||
692 LysAlaIleHisAlaAsnAlaValIleTyrLysPheSerProSerGlyAlaGlu 708

```



```

1183 AAACTCMAATGTCGCCAGTCAGTATGCCGTTATGAAATTTTGATGGT 1134
      ::::: |||
744 lavalserlyralpro.....Gly 750
1133 GGACCAACCGGTCAACAGTCATTTGCTATCATGTCAGCCAGTTTA 1084
      ||| ::::: |||
751 Gly.....ValLeuTyraIaValaInGlyLysProTyTy 762
1083 TCATTAATGACATGCGATTCGTGAACCGTGTGATCTTTCGCCGGTTC 1034
      | |||||
762 rGly.....AsperthrProValGInGlyPhe..... 771
1033 TCATTCCTGCTTGTGATGATGATGTAACGGTGTACTGTGAATTTCTA 984
      ::::: |||
772 .....MetIleasnPheserInGlyasp..... 779
983 AATGCTGATGATGCTGCTGTGAATTAATTTT.....CTAATTAATTT 940
      |||||
780 .....IleuValaspThrPheIleProAlaArgLysAsnPh 791
939 GAATATCCAAACAGATTTAATGCTGCGCAAGACCTCAGCTATTC 894
      :||| ||| ::::: |||
791 eGlnMetProHIsAspIleAlaIaGlyAspAspGlyThrValTy 806

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-540-804-12

seq_documentation_block:
Sequence 12, Application US/08540804
Patent No. 5919666
GENERAL INFORMATION:
APPLICANT: Young, Richard A.
APPLICANT: Koleske, Anthony J.
APPLICANT: Thompson, Craig M.
APPLICANT: Chao, David M.
TITLE OF INVENTION: No. 5919666el Factors Which Modify Gene
TITLE OF INVENTION: Transcription and Methods of Use Therefor
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/540, 804
FILING DATE: 11-OCT-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/521, 872
FILING DATE: 21-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/219, 265
FILING DATE: 25-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WHI94-03A2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-9540
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1226 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

```

US-08-540-804-12

alignment_scores:

Quality:	93.50	Length:	489
Ratio:	0.403	Gaps:	24
Percent Similarity:	47.444	Percent Identity:	19.836

alignment_block:

US-09-323-427-2/rev x US-08-540-804-12

Align seg 1/1 to: US-08-540-804-12 from: 1 to: 1226

```

1267 CATATGAGTAACTGCTTTTACATGGAAGCTGATAAAGCTTA..... 1223
      ::::: |||
279 TyraIaAspGlnAsnProLeuThrMetHisLysIleIleGlnLeuIle 295
1222 .....GTGCACAGATTGAG.....TATCTGAATCACAAAC 1192
      ::::: |||
295 uTrpSerIleHisProSerArgGlnPheAspHisTyrgIuSerInGln 312
1191 TGCTTTAAACTCAAAATTTGCCGATGCCAGTATGCCGTTATGAATTT 1142
      ||||| ||| |||
312 euValAlaIaLysLeuLeuLeuLeuArgIleAsn..... 322
1141 TCGATGTGAGCAACCGGTCAACCGATTCAAT.....TTGCT 1104
      ::::: |||
323 ...SerThrAspGluAspLeuHisGlnPheGlnIleGluAspAlaIleTr 338
1103 ATCATTTGTCACCCAGCTTTATCATTAATGACATCGATTTGAAACCGT 1054
      ||||| |||
338 pSerLeuValaPheGlnLeuAlaLysAsn..... 347
1053 TGATACTTTCGCGCGGTGTCGATTCCTGTTGCGATGATGTAACG 1004
      ::::: |||
348 .....PheSerAlaGlnLysArgValValSer 356
1003 GTGATACGTGGAATTTCAATGCTGATGATGCTGCTTGTATAATAT 954
      ::::: |||
357 TyrMetLeuProSerLeuTyraIrgLeuValAsnIleLeu..... 369
953 TTGCTAAATTAATTTGGAATATCCACAGATTTAATGGTCCCAAGACG 904
      ||| ::::: |||
370 .....IleThrTyrgIyIleIleLysValaProThrTyraIleArgLys 384
903 TCACGTAAT.....ACAAATATCGGATCGAT 878
      |||
384 euIleSerSergIyLeuLeuTyrgLysGlnAspSerAsnAspLysPheVal 400
877 CAC...AGCTTTTCTATCAATGCCAGATCAGTATTACCATTAAGAACA 831
      ||| ::::: |||
401 HisValGlnLeuLeuIleAsnLeuLysIleSerProLeuMetLysSerg 417
830 A.....ATACGAGATGTGTCGACCAACATGTTGACAGAACCA 793
      ::::: |||
417 nTyraSmetValLeuArgAsnValaMetGlnuTyraSpValaLysPheTyrg 434
792 AG.....GATTCGAGCTGTTAAACAGGTGTCGCCCAAAACCTG 749
      ::::: |||
434 IuIlePheAsnPhesaspGlnLeuValaGluIleThrGlnGlnIleLysMet 450
748 CTGACGTGCCCAACTTCGTTTACTCAAGAAAAGATCTGCGAACCAGG 699
      ||| |||
451 ArgIleLeuSerAsn.....AspIleThrAsnLeuG 461
698 AATATCATGTGATGATGACAACTGATTCACACCCCTGGAATTAAGAGATGA 649
      ::::: |||
461 nLeuSer.....LysThrProSerIleLysIleIem 472
648 TAATC.....AAGCTTGGCAGTTGATTTACGTCACG 617
      ::::: |||
472 etValaIaGlnuTrpTyrgLeuSerHisLeuCySerGlyIleLeuSerSer 488

```

```

616 GTGCAC.....TTTCACACATAT 597
489 ValAsnArgThrValLeuLeuLysIlePheLysIlePheCysIleAspLe 505
596 GGACACCTGTATACTTGTGCAGTACAAATGGAATCTGCATGTGCACC 547
505 UGValAlPheHisHisPhePheLysTrpIleGluPheIleValIleHisG 522
546 ATTGGCTTCTCAATGTTTATGGGTTTAAGCATTCATTCGTCGCCG 497
522 InLeuLeuSerAsp.....IleGluSerLeuGluAlaLeuMetAsp 535
496 TCATTATACATTCGTTTAATTCGTCCAA..ATCAGAAAGCAT.. 452
536 IleLeuLeuLysTrpGluLysLeuPheSerGluPheIleAsnAspHisI 552
451 .....AAAAATATGTAGATCATGACGAAACATATATAA.... 416
552 elePheThrLysThrPheIlePheIleTrpLysLysValLeuLysGluL 569
415 .....CTGCATATATATTCGTTCTTCTTATCATCCTCTTAATACTAA 371
569 ysAspValProAlaTrpAsnValThrSerPheMetProPheTrpLys... 584
370 TTTTAGCTACAAATATATAGTAT.....GTAGAAATATATTCG 330
585 PhePheMetLysAsnPheProPheValLeuLysValAspAsnAspLeuAr 601
329 TAATACATATAGTATATTTTC.....ATCAAACTCTCTCTA 292
601 gIleGluLeuGlnSerValTrpAsnAspGluLysLeuLysThrGluLysL 618
291 TCCTTTTATAGCTTCTGAAAACCTTATTCATTCATTCAGTAATC.... 248
618 euLysAsnAspLysSerGluValLeuLysValTrpSerMetIleAsnAsn 634
248 ..... 248
635 SerAsnGlnAlaValGlyGlnThrTrpAsnPheProGluValPheGlnVa 651
247 .....TTTATATGACATACTATTTGAATGTTTCATTCATTCAGGCC 208
651 lAsnIleArgPheLeuLeuHisAsnSerGluIleIleAspThrAsnHis 668
207 ATGAATAGTTCCTGTTGTATTCATTCATTCATTCACACTGTCTCTATTTA 158
668 eTrpSerGlnPheGlnLysAlaArgAsnAsnVal..... 678
157 TTCTAACAGTTTATCATTTGTGATATATACAAATTTATACCTTGATTCG 108
679 .....MetLeuLeuIleAlaThrAsnLeuLysGluTrpIleLysPheMe 693
107 CCCAATTTTATGGGATCATTCCTATTCGTGAACATTCACCTATTATT 58
693 tSerIlePheLeuLysArgLysAspPheThr...AsnLysAsnLeuIleG 709
57 GCATTATTCGCAATTAA 41
709 InLeuIleSerLeuLys 714

seq_name: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:US-08-218-265-12
seq_documentation_block:
: Sequence 12 Application US/08218265
: Patent No. 5922585
: GENERAL INFORMATION:
: APPLICANT: Young, Richard A.
: APPLICANT: Koleske, Anthony J.
: APPLICANT: Thompson, Craig M.
: TITLE OF INVENTION: No. 5922585el Factors Which Modify Gene
: NUMBER OF SEQUENCES: 35 Transcription and Methods of Use Thereof
: CORRESPONDENCE ADDRESS:

```

```

: ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
: STREET: Two Militia Drive
: CITY: Lexington
: STATE: MA
: COUNTRY: US
: ZIP: 02173
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/218,265
: FILING DATE: 25-MAR-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Granahan, Patricia
: REGISTRATION NUMBER: 32,227
: REFERENCE/DOCKET NUMBER: WH194-03
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-861-6240
: TELEFAX: 617-861-9540
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1226 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-218-265-12

```

```

alignment_scores:
  Quality: 93.50      Length: 489
  Ratio: 0.403       Gaps: 224
  Percent Similarity: 47.444   Percent Identity: 19.836

```

```

alignment_block:
US-09-323-427-2/rev x US-08-218-265-12 ..

```

```

Align seg 1/1 to: US-08-218-265-12 from: 1 to: 1226

```

```

1267 CATATGAGTACATGCTTTTACATGGAAGCTGATAAACAGTTA..... 1223
      ::::: ||| ::::: |||
279 TyrValAspGlnAsnProLeuThrMetHisLysIleIleGlnLeuIleLe 295
1222 .....GTGCACAGATTGAGG.....TATCGAATTCACAC 1192
295 utrpSerIleHisProSerArgGlnPheAspHisTrpGluSerAsnGlnL 312
1191 TGCTTTCAACTCAATATGTCCGATGCCAGTATGCCGTTATGAATTT 1142
      ||::: ||||| || || :::
312 euValAlaLysLeuLeuLeuArgIleAsn..... 322
1141 TGGATGTTGACCAACCGGTCAACAGTTCAAT.....TTGCT 1104
      ::::: ||::: ::::: ||:::
323 ...SerThrAspGlnAspLeuHisGlnPheGlnIleGluAspAlaIleTr 338
1103 ATCATGTGTACAGCAGTTTATCATATAATGACATGGCATTCGAAACCGT 1054
      ||||| ||| |||
338 pSerLeuValAlPheGlnLeuAlaLysAsn..... 347
1053 TGATACCTTTCTGCGCGGTTGTCCATTCCTGCTTTGTCATGTATGTAAG 1004
      :::: ||| ::: |||
348 .....PheSerAlaGlnLysArgValAlaSer 356
1003 GTGATATCTGTGAATTTCAATTCGATGATGATGTCCTTGATTAATTT 954
357 TyrMetMetProSerLeuTrpArgLeuLeuAsnIleLeu..... 369
953 TTGCTAAATATTTGGAATATCCACAGATTTAATGGCTGGCCAGGAAGC 904
      ||| ::::: ||| ::::: |||
370 .....IleThrTrpGlyIleIleLysValProThrTrpTrpIleArgLysL 384

```

```

903 TCACGAT.....ACAATATCGCAT 878
    |||
384 euIleSerSerGlyLeuLeuTyLeuGlnAspSerAsnAspLysPheVal 400
877 CAC...ACGTTTCATCATGCCAGATCAGTATTATTCATTAAACACA 831
    |||
401 HISValGlnLeuLeuIleAsnLeuLysIleSerProLeuMetLysSerI 417
830 A.....ATAGCGATGTGTGCACCAATGTTCAACACCA 793
    |||
417 nTyAsnMetValLeuArgAsnValMetGluTyAspValLysPheTyG 434
792 AG.....GATTCGGAGCTGTTAAACAGGTGTGCCGACCAAACTG 749
    |||
434 IuIlePheAsnPheAspGlnLeuValGluIlePheGluGlnIleLysMet 450
748 CTCGACGTGGCACTTCGTTTACTCAGAAAAGATCTGCAGAACCGGAG 699
    |||
451 ArgIleLeuSerAsn.....AspIleThrAsnLeuG1 461
698 AATATCATTTGATGATGACACTGATATCAACACCCCTGAATTAGCATGA 649
    |||
461 nLeuSer.....LysThrProLeuSerIleLysIleM 472
648 TAATC.....AAGCTTCCAGTTGATTATTCAGTCACC 617
    |||
472 etValAlaGluIrrTyLeuSerHisLeuLysSerGlyIleLeuSerSer 488
616 GTGCAC.....TTTCACAACATAAT 597
    |||
489 ValAsnArgThrValLeuLeuLysIlePheLysIlePheCysIleAspLe 505
596 GGACAACTGTAATACCTGCTGCAGTACAAATGSAATGATGATGATGACC 547
    |||
505 uGluValPheHisHisPhePheLysTrpIleGluPheIleValIrrHisG 522
546 ATTGGCTTTCATGATTATGCGTTAAGCATTCATGATTGCTGCGC 497
    |||
522 InLeuLeuSerAsp.....IleGluSerLeuGlnAlaLeuMetAsp 535
496 TCATTATTACATTTTCGTTAAATTCGTCGAA...ATCAGAAAGCAT 452
    |||
536 IleLeuLeuCysTyrrGlnLysLeuPheSerGlnPheIleAsnAspIleI 552
451 .....AAAATATGTAGATCATCGACGACATATATATAA... 416
    |||
552 euLeuPheThrLysThrPheIlePheIleTyrrLysValLeuLysGlu 569
415 .....CTGCCATATATTCGTTCTTCTTATCATCCTCTTAATTAACATA 371
    |||
569 yAspValProAlaTyrrAsnValThrSerPheMetProPheTrpLys... 584
370 TTTTAGCTACAAATATATAGTAT.....GTAGAAATATATATCTG 330
    |||
585 PhePheMetLysAsnPheProPheValLeuLysValAspAsnAspLeuAr 601
601 gIleGluLeuGlnSerValTyrrAsnAspGluLysLeuLysThrGluLysL 618
291 TCGCTTTTATGCTTCGAAAGCTTATTCATTTATTCAGTAATC..... 248
    |||
618 euLysAsnAspLysSerGluValLeuLysValTyrrSerMetIleAsnAsn 634
248 ..... 248
635 SerAsnGlnAlaValGlyGlnThrTyrrAsnPheProGluValPheGluVa 651
247 .....TTTATATGATACTATGTTAAATGTTATCATCATATGAGCC 208
    |||
651 LAsnIleArgPheLeuLeuHisAsnSerGluIleIleAspThrAsnThr 668
207 ATGAATAGTTCGTTGTTATATATCATCATATCAACTGTGCTATTTTAA 158

```

```

668 erLysGlnPheGlnLysAlaArgAsnAsnVal..... 678
157 TTTACAGATTATCATTTGATATATATACAAATTAATTAACCTGTATG 108
679 .....MetLeuLeuIleAlaThrAsnLeuLysGluTyrrIleLysPheMe 693
107 CCCAATTTTATGCGCATCATTTCTATTCGTAAACATTCATCATTTT 58
    |||
693 tSerIlePheLeuLysArgLysAspPheThr...AsnLysAsnLeuIleG 709
57 GCATTATTCGCAATTAA 41
    |||
709 InLeuIleSerLeuLys 714

seq_name: /cgn2_6/ptodata/1/iaa/5A-COMB.pep:US-07-906-349A-6

seq_documentation_block:
; Sequence 6, Application US/07906349A
; Patent No. 5434064
; GENERAL INFORMATION:
; APPLICANT: Schlessinger, Joseph
; APPLICANT: Skolnik, Edward Y.
; TITLE OF INVENTION: A NOVEL EXPRESSION-CLONING METHOD FOR
; TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE KINASES
; TITLE OF INVENTION: TARGET PROTEINS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brodsky and Nelmark
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/906,349A
; FILING DATE: 30-JUN-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/643,237
; FILING DATE: 18-JAN-1991
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 801 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-906-349A-6

alignment_scores:
    Quality: 93.00      Length: 77
    Ratio: 3.000        Gaps: 1
Percent Similarity: 40.260      Percent Identity: 28.571

alignment_block:
US-09-323-427-2/rev x US-07-906-349A-6  ..
Align seg 1/1 to: US-07-906-349A-6 from: 1 to: 801

798 ACCACAAGATTGCGAGCTGTTAAACAGGTGTGCCGACGACAAACCTG 749
|||||
573 thrThrAlaAlaIleThrCysAlaCysThrGlyCysThrGlyCysThrCys 589

```



```

748 CTGAGCTGGCGCACTGCTTACTCAAGAAAAGATCGAAGACCGAG 699
||||:||||:||||:
589 scythrCysThrThrThrCysCysCysThrCysCysThrThr... 604
698 AATATCATGTGATGACGAAGTATATCAACACCGTTGAATTAGCATGA 649
604 ..... 604
648 TAATCAAGCTTGGCAGTGTGATTACGTCACCGCTGCTGCAACATA 599
||||| 613
605 .....ThrGlyThrCysThrThrThrThr 613
598 ATGCACAACCTGTATAGTCTGCTGCAAGTACA 568
613 rThrThrThrCysAlaThrCysCysThrThr 623

seq_name: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:us-08-415-751-35

seq_documentation_block:
; Sequence 35, Application US/08415751
; Patent No. 5643772
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: POLYPEPTIDES BINDING ANTI-
; TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA
; TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID
; TITLE OF INVENTION: VECTOR AND TRANSFORMED HOST AND
; TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PHILLIPS, MOORE, LEMPFO & FINLEY
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: California
; COUNTRY: United States of America
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Kb storage
; COMPUTER: PC
; OPERATING SYSTEM: DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/415,751
; FILING DATE: 03-Apr-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/071,880
; FILING DATE: June 1, 1993
; APPLICATION NUMBER: 07/891,301
; FILING DATE: May 29, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hana Dolezalova
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.19-2 (HMD)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-1677
; TELEFAX: (415) 324-1678
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 362 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Cryptosporidium parvum
; FEATURE:
; NAME/KEY: Positions coded by nonsense codons are

```

```

; NAME/KEY: Identified as Xaa.
; US-08-415-751-35

alignment_scores:
Quality: 91.50 Length: 151
Ratio: 1.220 Gaps: 9
Percent Similarity: 49.669 Percent Identity: 25.828

alignment_block:
US-09-323-427-2/rev x US-08-415-751-35 ..

Align seg 1/1 to: US-08-415-751-35 from: 1 to: 362

853 TCAGTATATACCATTAAGAACCAATAGACGAATGTGTGCGCAC...AA 807
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
20 SerAsnLeuLeuLeuGlnHisSerArgArgHisPhePheGluValGlu 36
806 TGTTTCAGAACCAAGAGATTTCGAGCTGTAAACAGGTGGTCCGAGC 757
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
36 yMeLArgAsn***ProHisArgGluLeuHisGluLeuValValLysG 53
756 AAACCTGTGCGACCTGCGCAC...TTTC 731
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
53 LrnrIle***HisLeuValSAsnArgCysLysTyrGlnSerThrHis 69
730 GTTACTCAAGAAAAGAT...CTGCAGAACCGAGAAATATCATGTAT 687
:: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
70 LeuGlnSerGlnIleAspPheLysLeuGlnHisAsnArgLeu... 82
686 GTACGAACGTGATATCAACACCGCTGAATATTAGCGATGATATCAAGCTTT 637
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
83 .....LeuValGlu****LeuGlnLeuArgLeuValValLeuLeu* 97
636 GCCAGTGTATTACGTCACCGCTGCTGCAACATATAGGACA... 592
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
97 *****LeuGlnPhe.AspleuCysValLeu*****Trp**ArgLe 113
591 .ACCTGTATATACCTGCTGACGTACCAAAATGAAT... 559
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
113 uCysCysCysGlyCysCysSerLysIleTrpAspAsnCysCysSerL 130
558 .....CTGCATGTACCATTTGGCTTCTC 535
130 eudheTrpCysCysCysCysCysLeuAsnTrpThrLeuTrpTrp... 145
534 AATGTTATGGTTTAAAGCATGATGATGATGCGCGCATATATACCA 485
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
146 .....TyrGly.....CysCysCysCysCys***LeuTrpTyrAs 157
484 T 484
157 n 157

seq_name: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:us-08-303-238-3

seq_documentation_block:
; Sequence 3, Application US/08303238
; Patent No. 5654270
; GENERAL INFORMATION:
; APPLICANT: RUOSLAHTI, ERKKI I.
; APPLICANT: LONGAKER, MICHAEL T.
; APPLICANT: WHITBY, DAVID J.
; APPLICANT: HARPER, JOHN R.
; APPLICANT: PIERSCHACHER, MICHAEL D.
; APPLICANT: BORDER, WAYNE A.
; TITLE OF INVENTION: INHIBITORS OF CELL REGULATORY FACTORS
; TITLE OF INVENTION: AND METHODS FOR PREVENTING OR REDUCING SCARRING
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
; CITY: SAN DIEGO

```

```

STATE: CALIFORNIA
COUNTRY: UNITED STATES
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/303,238
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/978,931
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: KONSKI, ANTOINETTE F.
REGISTRATION NUMBER: 34,202
REFERENCE/DOCKET NUMBER: P-LA 9453
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-535-9001
TELEFAX: 619-535-8949
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 368 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-303-238-3

alignment_scores:
Quality: 91.50 Length: 299
Ratio: 0.738 Gaps: 14
Percent Similarity: 41.472 Percent Identity: 19.732

alignment_block:
US-09-323-427-2/rev x US-08-303-238-3 ..
Align seg 1/1 to: US-08-303-238-3 from: 1 to: 368

775 AACAGGTGTCGCCGACGAAACCTGCTGCGACCTGCGCAACTTCGTTTA 726
|||||:|||||:|||||:|||||:|||||:
86 LysGluIleSerProAspThrLeuLeuAspLeuGlnAsnAspIle 102
725 CTCGCA.....AGAAAAGAT..... 713
|||||:|||||:|||||:|||||:|||||:
102 eserGluLeuArgLysAspAspPheLysGlyLeuGlnHisLeuTyrAla 119
712 .....CTGCGAACCGAGAAATATCATTTGATGAGAACTGATATCAACA 668
|||||:|||||:|||||:|||||:|||||:
119 euValLeuValAsnAsnLysIleSerLysIleHisGluLysAlaPheSer 135
667 CCCTTGAATTAAGCATGATATATCAAGCTTTGCCAGTTGATTTACGTCAC 618
|||||:|||||:|||||:|||||:|||||:
136 ProLeuArgAsnVal.....GlnLysLeuTyrIleSe 146
617 CGTGCACCTTGCACACATATGAGACAACTGTATACTTGTCTGCAGTACA 568
146 rLys..... 147
567 AAATGGAATCTGCATGTCACCATTTGGCTTCTCAATGTTTATGGGTTTAA 518
|||||:|||||:|||||:|||||:
148 .....AsnHisLeuValGlu..... 152
517 GCATTCATTCATTCGTCGCCGACGAAACCTGCTGCGACCTGCGCAACTTCGTTTA 468
|||||:|||||:|||||:|||||:|||||:
153 .....IleProAsnLeuProSerSerLeu.....Va 162
467 CCAATTCAGAGGCAATAAATATATGTTTGA.....ATCA 433
|||||:|||||:|||||:|||||:|||||:
162 GluLeuArgIleHisAspAsnArgIleArgLysValProAsnGlyValP 179

```

```

432 TCGAAGCAATA..... 422
|||||:|||||:|||||:|||||:|||||:
179 heserGlyLeuArgAsnMetAsnCysIleGluMetGlyGlyAsnProLeu 195
421 .....ATAAAGTCCCAFA 408
|||||:|||||:|||||:|||||:|||||:
196 GluAsnSerGlyPheGluProGlyAlaPheAspGlyLeuLysLeuAsnTyr 212
407 TATATTCGTTTCCTTATTCATCTCTTAATACTAATTTTATAGCAACA 358
|||||:|||||:|||||:|||||:|||||:
212 rLeuArgIleSerGluAlaLys.....LeuThrng 222
357 ATATATAGTATAGTAAGAAATATATCTGTAATACAAATAGTATATTTTC 308
|||||:|||||:|||||:|||||:|||||:
222 rLysLeu.....ProLysAspLeuProGluThrLeuAsnGlnLeuHis 235
307 ATCAAAACCTTCCTATTCGCTTTTATAGCTTGAAGAACTATTCATTA 258
|||||:|||||:|||||:|||||:|||||:
236 LeuAspHisAsnLysIleGlnAlaIleGluLeuGlnAspLeuLeuArgTyr 252
257 TTCAGTAATCTTTTATATGCATATCTATTGTAATGTTTCATCATTAAGGCC 208
|||||:|||||:|||||:|||||:|||||:
252 rSerLysLeuTyrArgLeu.....GlyLeuGlyH 262
207 ATGAATATGTTTCGTTTGTATATCA...TCATTAATCACTTGTCT... 164
|||||:|||||:|||||:|||||:|||||:
262 IsasnglnIleArgMetIleGluAsnGlySerLeuSerPheLeuProThr 278
164 ..... 164
279 LeuArgLysLeuHisLeuAspAsnAsnLysLeuAlaArgValProSerG 295
163 .....ATTATATCTAACAGTTTATCATTTGTGATATA 130
|||||:|||||:|||||:|||||:|||||:
295 yLeuProAspLeuLysLeuLeuGlnValTyrIleHisSerAsnAsnI 312
129 TCACAAATTATACCTTG.....TATTGCCCAATTTTATGCGCATC 89
|||||:|||||:|||||:|||||:|||||:
312 lerThrLysValGlyValAsnAspPheCysPrometGlyPheGlyVal 327

seq_name: /cgn2_6/prodata/1/1aa/backfilese1.dep:5340934-2
seq_documentation_block:
Patent No. 5340934
APPLICANT: TERMINE, JOHN D.; YOUNG, MARIAN F.; FISHER, LARRY W.
ROBEY, PAMELA G.
TITLE OF INVENTION: CDNA SEQUENCES OF HUMAN BONE MATRIX PROTEINS
NUMBER OF SEQUENCES: 13
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/432,044
FILING DATE: 03-NOV-1989
SEQ ID NO: 2
LENGTH: 368
5340934-2

alignment_scores:
Quality: 90.50 Length: 299
Ratio: 0.730 Gaps: 14
Percent Similarity: 41.472 Percent Identity: 19.732

alignment_block:
US-09-323-427-2/rev x 5340934-2 ..
Align seg 1/1 to: 5340934-2 from: 1 to: 368

775 AACAGGTGTCGCCGACGAAACCTGCTGCGACCTGCGCAACTTCGTTTA 726
|||||:|||||:|||||:|||||:|||||:
86 LysGluIleSerProAspThrLeuLeuAspLeuGlnAsnAspIle 102
725 CTCGCA.....AGAAAAGAT..... 713
|||||:|||||:|||||:|||||:|||||:
102 eserGluLeuArgLysAspAspPheLysGlyLeuGlnHisLeuTyrAla 119

```

```

712 .....CTGAGAACCGGAGATATCATTTAGTACAGACTGATATCACA 668
    ||| |||.....||| ||| .....|||
119 euValleuValasnsnlyslleerlysllehsnglyslalaPhseSer 135
667 CCCTTGAAATTAACGATGATATACACCTTTGCCAGTTGATTTACGTGAC 618
    ||| |||.....||| ||| .....|||
136 ProleuArgnsnVal.....GlnlyseuYrilese 146
617 CGGCACTTGTGCAACATATGACAACTGTAACTCTGTCAGTACA 568
146 rlys..... 147
567 AATGGAATGTCATGTCACCATTTGGCTTCTCAATGTTATGGGTTTA 518
    ||| |||.....||| ||| .....|||
148 .....AsnHslleuValglu..... 152
517 GCATTCGATTCGATTCGCTGCGTCATATTAACCATTTGTTAAATTCGT 468
    ||| |||.....||| ||| .....|||
153 .....IleProAsnleuProSerSerleu.....Va 162
467 CCAATCAGAGGACATAAAAATATGTTGA.....ATCA 433
    ||| |||.....||| ||| .....|||
162 lAspValArgyleHlsAspAsnArgyleArglyslValProlyslYvalP 179
432 TCGAGCACAATA..... 422
179 heserGlyleuArgasMetasnGlyleGlnMetGlyslasProleu 195
421 .....ATAAAGTGCACATA 408
196 GluAsnSerGlyPheGlnProGlyAlaPheAspGlyleuLysleuAsnTy 212
407 TATATTCGCTTCTTATCATCCTCTTAAATACATAATTTAGCTAACAA 358
    ||| |||.....||| ||| .....|||
212 rleuArgyleSerGluAlaLys.....LeuThrG 222
357 ATATATGATGATAGGAATTAATTAATACATATGATATTTTC 308
    ||| |||.....||| ||| .....|||
222 lYle.....ProlysAspLeuProGlnThrleuAsnGlyleuHls 235
307 ATCAAAACTTCTTATACGCTTTTATAGCTTCTGAAAGCTTATTCATTA 258
    ||| |||.....||| ||| .....|||
236 LeuAspHlsAsnlyslleGlnAlaileGlyleuGlnAspLeuLeuArgTy 252
257 TTCAGAAATCTTTATATGCATACATATTTAATGTTTCATCATTAAGGCC 208
    ||| |||.....||| ||| .....|||
252 rserlyseuTyArgleu.....GlyleuGlyH 262
207 ATGATAGTATTCGTTGTTATATCA...TCATATCAACTGTGCT... 164
    ||| |||.....||| ||| .....|||
262 lAsnGlnlleArgMetileGlnAsnGlySerleuSerPheleuProThr 278
164 ..... 164
279 LeuArgGlyleuHlsleuAspAsnAsnlyslleuAlaArgValProSerG 295
163 .....ATTTATTTTAAACAGTTTATCATTTGTTGAATAA 130
    ||| |||.....||| ||| .....|||
295 yleuProAspLeuLysleuGlnValValTyrlleuHlsSerAsnAsn 312
129 TCACAAATTTATACCTTG.....TATGCCCAATTTTATAGGCATC 89
    ||| |||.....||| ||| .....|||
312 lethrlyslValGlyValAsnAspPheCysProMetGlyPheGlyVal 327
seq.name: /cgn2.6/ptodata/1/aa/5A_COMB.pep.us-08-325-071-67
seq_documentation_block:
; Sequence 67, Application US/08325071
; Patent No. 5587311
; GENERAL INFORMATION:
; APPLICANT: COBON, Stewart Gary
; APPLICANT: MOORE, Joanna Terry

```

```

APPLICANT: JOHNSON, Law Anthony York
APPLICANT: WILDSEN, Peter
APPLICANT: KEMP, David Harold
APPLICANT: SRISKANTHA, Alagacone
APPLICANT: RIDING, George Alfred
APPLICANT: RAND, Keith No. 3387311man
TITLE OF INVENTION: DNA Encoding A Cell Membrane
TITLE OF INVENTION: Glycoprotein Of A Tick Gut
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESS: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/325,071
FILING DATE: 14-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,109
FILING DATE: 17-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/926,368
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/242,196
FILING DATE: 06-JUL-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU87/00401
FILING DATE: 27-NOV-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P14912
FILING DATE: 16-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P12570
FILING DATE: 19-JUN-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P9196
FILING DATE: 27-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 60042/111 BIAU
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SRO ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 650 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-325-071-67

alignment_scores:
Quality: 90.00 Length: 359
Ratio: 0.612 Gaps: 22
Percent Similarity: 40.947 Percent Identity: 21.727

alignment_block:
US-09-323-427-2/rev x US-08-325-071-67 ..
Align seg 1/1 to: US-08-325-071-67 from: 1 to: 650
1476 TGCATTCGAGACATGT.....TTATG 1454
|||||

```

147 CysValProThrThrCysLeuArgProAspLeuThrCysLysAspLeu 163
1453 TGAAGGCTCTTATGATCAAGAAGTTGCCGTAATGATGAAGGTCAGT 1404
163 sGluLysAsnLeuLeuGlnArgAsp.....Sera 173
1403 CAAGTCCCGGAATTCCTCCATTGATTCATGCAATGTTGCCGCTAC 1334
173 rGcysCysGlnGlyTyrPasnThr.....AlaAsnCysSerAla1a 186
1353 ACGATCTCTGAATCCACGCTGATTTTGTAC..... 1321
187 ProProAlaAspSer.....TyrCysSerProGlySerProLysG1 200
1320AACACTGCTGTCAT.....TTCG 1302
200 yProAspGlyGlnCysLysAsnAlaCysArgThrLysGlnAlaGlyPheV 217
1301 TTTTCATTCATTATTGTTACCAAGTTGATCGTCATATCGATCAATG 1252
217 aLcysLysHisGlyCys.ArgSerThrAspLysAlaTyrGlyCysThrCy 233
1251 C.....TTTACATGGAAGCTGATAAACAGTTAGTCACAGATG 1211
233 sProSerGlySerThrValAlaGlnAspGlyLethrCysLysSerIles 250
1210 AGGTATCTGAATCAACACTGCTTTCACAACTCAATGTC.....CCGATG 1164
250 eTyrThrValSerCysThrValGlnGlnLysGlnThrCysArgProThr 266
1163 CCAATGACGCTATGAA.....ATTTGAGTGGTGACCAACCG 1123
267 GlnAspCysArgValGlnLysGlyThrValLeuGlnCysGlnCysProThr 283
1122 TCAACAGATTCAATTGCTATCATTTGCTCAGCCAGTTATCATTAATGA 1073
283 glnL.....HisLeuValGlyAsp.....T 290
1072 CAGCCGATTCTGAACGCTGATCTTTCGCGGTTGCCATTCCTGC 1023
290 hrcysIleSerAspCysValAspLysCys.....HisGlnGlu 303
1022 TTTCGATGATGCT..... 1008
304 PheMetAspCysGlyValTyrMetAsnArgLnsrCysTyrCysProTr 320
1007AACGTAFACTGCTGAATTTCTAATGCTGATGAT 971
320 pLysSerArgLysProGlyProAsnVal.....AsnIleAsnGluC 334
970 GTGCTCTGATTAATATTGCTAAT.....AATTTGGA 936
334 yLeuLeuAsnGlnTyrTyrThrValSerPheThrProAsnIleSer 350
935 TANCAACAGATTATAGCTGCGCAAGAAGCTCAGCTTACAATATGC 886
351 PheAspSerAsp.....HisCysLysArgTyrG1 360
885 GGATTCGATCAGACGCTTTCTATCATATGCCAGATCAGTATTCATTAAG 836
360 uAspArg.....ValLeuGlnAlaIleArgThrSerIleGlyLysGluV 375
835 AACCAATAGCAATGTTGCGACCAATGTTCAAGACCAAGATTC 786
375 aLpHeLysValGlnIleLeu.....AsnCysThrGlnAsp..... 386
785 GAGAGCTTAACAGAGTGGTGGCAGCAAAACCTCTGCAGCTGGCGCA 736
387IleLysAlaArgLeuIleAlaGlnLysProLeuSerLysTyrVa 401
735 ACTTCGT..... 729
401 lLeuArgLysLeuGlnAlaCysGlnHisProIleGlyGluTyrCysMet 418

728TTACTCAGAAAAGATCTGCA.....GAACCG 702
418 eTyrProLysLeuLeuIleLysLysAsnSerAlaThrGlnIleGlnGlu 434
701 GAGATATCATTTGATGATGACAGACTGATATCAACACCTTGAATTAAGCA 652
435 GlnAsnLeuCysAspSerLeuLeuLysAsnGlnGlnAlaIleTyrLysG1 451
651 TGATATCAAGCTTTGCCAGTTGAT 627
451 yGlnAsnLysCysValLysValAsp 459
seq_name: /cgn2.6/ptodata/1/iaa/5B_COMB.pep:US-08-477-451-22

seq_documentation_block:
Sequence 22, Application us/08477451
Patent No. 5928865
GENERAL INFORMATION:
APPLICANT: COVACI, Antonello
TITLE OF INVENTION: Helicobacter pylori CagI Region
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESS: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 485 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-451-22

alignment_scores:
Quality: 89.50 Length: 564
Ratio: 0.393 Gaps: 28
Percent Similarity: 40.426 Percent Identity: 18.085

alignment_block:
US-09-323-427-2/rev x US-08-477-451-22 ..
Align seg 1/1. to: US-08-477-451-22 from: 1 to: 485

1447 GTCTTATGATCAAGAAGTTGCCGTAATGATGAAGGTGAGCTCAAGTT 1398
10 ValLeuValTyrProIleLeuPheLeuPheAlaLeuIleLys.. 25
1397 GCCGGAATTCATCTTCATTTGATTCATGCAATGTTGCCGCTA..... 1355
26 ProSerPhePheTyrTyrThrThrTyrLeuLeuLeuValSerLeus 42

370 yGlnAsnLysCysValLysValasp 378

seq_name: /cgn2.6/ptodata/1/iaa/5A.COMB.pep:US-08-325-071-63

seq_documentation_block:
; Sequence 63, Application US/08325071
; Patent No. 5587311

GENERAL INFORMATION:

APPLICANT: COBON, Stewart Gary
APPLICANT: MOORE, Joanna Terry
APPLICANT: JOHNSON, Law Anthony York
APPLICANT: WILLADSEN, Peter
APPLICANT: KEMP, David Harold
APPLICANT: SRISKANTHA, Alagacone
APPLICANT: RIDING, George Alfred
APPLICANT: RAND, Keith No. 5587311man
TITLE OF INVENTION: DNA Encoding A Cell Membrane
TITLE OF INVENTION: Glycoprotein Of A Tick Gut
NUMBER OF SEQUENCES: 71

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/325,071

FILING DATE: 14-OCT-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/062,109

FILING DATE: 17-MAY-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/926,368

FILING DATE: 07-AUG-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/242,196

FILING DATE: 06-JUL-1988

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU87/00401

FILING DATE: 27-NOV-1987

PRIOR APPLICATION DATA:

APPLICATION NUMBER: AU P14912

FILING DATE: 16-OCT-1987

PRIOR APPLICATION DATA:

APPLICATION NUMBER: AU P12570

FILING DATE: 19-JUN-1987

PRIOR APPLICATION DATA:

APPLICATION NUMBER: AU PH9196

FILING DATE: 27-NOV-1986

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A
REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 60042/111 BIAU

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202 672 5300

TELEFAX: 202 672 5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 63:

SEQUENCE CHARACTERISTICS:

LENGTH: 650 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-325-071-63

alignment_scores:

Quality: 89.00 Length: 359
Ratio: 0.614 Gaps: 22
Percent Similarity: 40.390 Percent Identity: 22.284

alignment_block:

US-09-323-427-2/rev x US-08-325-071-63

Align seq 1/1 to: US-08-325-071-63 from: 1 to: 650

```

1476 TGCATCGAAGACATGT.....TTATG 1454
||||| |||||
147 CysValProThrThrcysLeuArgProAspLeuThrcysLysAspLeu 163
||||| |||||
1453 TGAAGGCTTTATGATCAAGAAGGTCGCGTATGATGAAGGTGACG 1404
||||| |||||
163 sglLysAsnLeuLeuGlnArgsp.....Ser 173
||||| |||||
1403 CAAGTTCGCGGAATTCACCTTCATTTGATTCATGCAAGTTCGCGTAC 1354
||||| |||||
173 rgcyscysglnGlyTrpAsnThr.....AlaAsncysSerAlaIala 186
||||| |||||
1353 AGCATCTCGAATCCACGCGGTATTTTGTACAC..... 1318
||||| |||||
187 ProProAlaAspser.....TyrCysSerProGlySerProLysgl 200
||||| |||||
1317 .....ACTGTGTGAT.....TTGCG 1302
||||| |||||
200 yProAspGlyGlnCysIleAsnIlaCysLysMetLysGlnAlaGlyPhe 217
||||| |||||
1301 TTTCATCCATTTATTTGTTCACAAAGTTCGATCGATATCGAGTACATG 1252
||||| |||||
217 alCysgluHnsgLysCys.ArgSerThrAlaLysAlaTyrlGlnCysThr 233
||||| |||||
1251 C.....TTTACATGCAAGCTGTAAACAGTTATGCAACATG 1211
||||| |||||
233 sProArgGlyPheThrValAlaGlnAspGlyIleThrCysLysSerIle 250
||||| |||||
1210 AGGTATCTGAATTCACACACTGCTTTCAACTCAATTCGTC...CGATG 1164
||||| |||||
250 erHsThrValSerCysThrIlaGlnGlnLysglThrCysArgProThr 266
||||| |||||
1163 CCAGTATGCCGTTATGAA.....ATTTGATGTGTGACCAACCGG 1123
||||| |||||
267 GluAspCysArgValHsLysglYthrValLeuCysGlnCysArgProThr 283
||||| |||||
1122 TCACACAGTTCAATTTGCTATCATGTCGACGACGTTATTCATTAATGA 1073
||||| |||||
283 nGln.....HsLeuValGlyAsp.....T 290
||||| |||||
1072 CATGCGATTCGAACCGTTGATACCTTCGCGGTTGTCATTCCCTGCG 1023
||||| |||||
290 hrcysIleSerAspCysValAspLysLysCys.....HsGlnGlu 303
||||| |||||
1022 TTGTGCGATGATGCT..... 1008
||||| |||||
304 PheMetAspCysGlyValYlmeLsnArgGlnSerCysTyrCysProTr 320
||||| |||||
1007 .....AACGGTATCGTGTGAATTTCAATTAAGTGTGATGAT 971
||||| |||||
320 pLysSerArgLysProGlyProAsnVal.....AsnIleAsnGlyC 334
||||| |||||
970 GTGCTCTTGATAAATATTTGCTAAT.....AATTTGAA 936
||||| |||||
334 yLeuLeuAsnGlnLysTrpYlThrIleValSerPheThrProAsnIleSer 350
||||| |||||
935 TATCCAACAGATTTAATGCTGCGCAAGACCTCAGTATACAAATATGCG 886
||||| |||||
351 PheAspSerAsp.....HsCysLysTrpTyrG 360
||||| |||||
885 GGATGATCACAGCTTTTCTATCAATGCGACATCGATTAACATTAAAG 836
||||| |||||
360 uAspArg.....ValLeuGlnAlaIleArgThrSerIleGlyLysGlu 375

```

```

835 AACCAATTAGCGAATGTGTGACCAATGTTGAGAACCAAGATTC 786
    :: ||| :: ::|||::
375 alPheLysValGluIleLeu.....AsnCysThrGlnAsp..... 386
    ::|||::
785 GGACCTGTTAAACAGGTGTGCCGACAAACCTGCTGCAGCTGCCGA 736
    ::|||::
387 .....IleLysAlaArgLeuIleAlaGluLysProLeuSerAsnHisVa 401
    ::|||::
735 ACTTCGT..... 729
    |||||
401 IleuArgLysLeuGlnAlaLacysGluHisProIleGlyGluTrpCysMetM 418
    |||||
728 .....TTACTCAAGAAAGATCTGCA.....GAACCG 702
    |||||
418 eTyrProLysLeuLeuIleLysLysAsnSerAlaThrGluIleGluGln 434
    |||||
701 GAGATATCATGTGATGTACGAACTGATATCAACACCCCTTGAATTAGCGA 652
    |||||
435 GluAsnLeuCysAspSerLeuLeuLysAsnGlnGluAlaAlaTyrLysG1 451
    ::|||
651 TGATATCAAGCTTTGCCAGTTGAT 627
    ::|||
451 yGlnAsnLysCysValLysValAsp 459

```



```

A:Residues: 1-308 <SEB>
A:Cross-references: EMBL:M55997; NID:g15627L; PID:g156272
C:Genetics:
A:Gene: CUT-1
A: Introns: 245/3

alignment_scores:
    Quality: 897.50      Length: 302
    Ratio: 3.989        Gaps: 5
    Percent Similarity: 74.503      Percent Identity: 60.596

alignment_block:
    US-09-323-427-2/rev x S27799 ..

Align seg 1/1 to: S27799 from: 1 to: 308

1253 TGCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
6 CyspHeTyMeGlueAspIserAspIsthrValIserThrgInIleGluValSe 22
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
1203 TGAATACACAAGCTGCTTTTCAAACTCAAAATTTGCCGATGCGAGTATGCC 1154
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
22 rsplseuThrThrAlaPheGlnThrGlnValValProMetProValCysL 39
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
1153 GTTATGAATTTTGGATGGTGGACCAACCGGTCAACAGTTCATTTGCT 1104
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
39 ystYrGluIleuAspGlyGlyProserGlyGlnProIleGlnHea 55
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
1103 ATCATTTGGTCAGCAGCTTTATCATTAATGACATCGATTCGAAACGCT 1054
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
56 ThrIleGlyGlnGlnValIleThyHisIleYstYrPheCysAspSerGlnThr 72
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
1053 TGATACCTTTCGCGCGGTTCGCCATTCCTGCTTTCGATGATGATACG 1004
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
72 rsAPThrPheCysAlaValAlaHisSerCysThrValAspAspGlyAsnG 89
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
1003 GTGATACGTGGAAATTTCAATGCTGATGATGTCCTTGTATTAATAT 954
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
89 LysAPThrValGlnIleuAsnGlnGlnGlyCysAlaIleuAspIysPhe 105
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
953 TTGCAATAAATTTTGGAAATATCCACACAGATTTAATGGCTGGCCAGAAGC 904
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
106 LeuIleuAsnAspLeuGlnIleuProThrAspLeuMetAlaGlyGlnGluAl 122
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
903 TCACGTATACAAATATGCGGATCGATCGACACAGCTTTTCTATCAATGCCAGA 854
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
122 ahIstValIleYstYrAlaAspArgSerGlnLeuPheTyGlnCysGlnI 139
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
853 TCAGTATTACCATTAAGAACAATATGCAATGTGTTGCACCAATATGT 804
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
139 IeSerIleThrIleIleYstYrAlaAspArgSerGlnLeuPheTyGlnCysGlnI 155
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
803 TCAGAACCCACAGGATTCGAGAGCTTTAAACACAGGTGCGCCGACAA 754
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
156 SerGlnProGlnGlyIleuAsnGlnGlyIleuValIleGlnAlaGlyAl 172
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
753 ACCTGCTGCAGCT..... 741
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
172 ahIstAlaAlaIleAlaIleProGlnAlaGlyValGlnGluValGlnAlaIle 189
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
741 ..... 741
189 rovalGlyAlaIleAlaProValAlaAlaProValAlaAlaIleAlaIle 205
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
740 ..... GCGCAACTTCGTTTACTC...AAGAA 718
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
206 ProAlaValProAlaIleThrIleuAlaGlnLeuArgLeuIleuArgIleYst 222
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
717 AAGATCT...CGAGAACCAGGAAATATCATTTGATGTACGACACGATATCA 677
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204
222 sarGerPheGlyGluAsnGlnGlyIleuAspValArgValGlnIleu 233
|||||TTCCTTTCATGGAAGCATATAAACAAGTTAGTGCACAGATGAGGATC 1204

```



```

A:Residues: 1122 <GAR>
A:Cross-references: GB:AE001417; GB:AE001362; NID:93845271; PID:93845273; TIGR:PF00770C
A:Experimental source: Clone 3D7
A:Genetics:
A:Gene: PFB0770C

alignment_scores:
      Quality: 116.50      Length: 508
      Ratio: 0.492      Gaps: 31
      Percent Similarity: 46.654      Percent Identity: 21.654

alignment_block:
05-09-323-427-rev x F71606 ..

Align seg 1/1 to: F71606 from: 1 to: 1122

1404 TCAAGTTCCGCGAATTTCACTTCCATTGATTCATGCAAGTTCGGCGTA 1355
      ::::::::::::::: |||
504 AsnThrCysAsnAsnHisThr.....CysAsnAsnHis 514
      ::::: |||
1354 CAGCACTCTGCATCCACGCTGATTTTGTACACACCACTGTTGTCATT 1305
      ::::: |||
514 sThrCysAsnAsnHisThr.....CysAsnAsnHisThrCysAsn 527
      ::::: |||
1304 TCGTTTCATCCATATTGTTTACCAAGTGTAGTCGATTCGAGTACA 1255
      ::::: |||
528 .....AsnHisThrCysAsnAsnHisThr 535
      ::::: |||
1254 ATGCTTTTACATGGAAGCTGATTAACACGTTAGTGCACAGATTGGATG 1205
      ||| |||
536 CysAsnAsnHisThrSer...AspAsnAsnThrCysAsn..... 547
      ::::: |||
1204 CTGAATACACACTGCTTTTCAACTCAAAATGTCGCCGATGCCAGTATGC 1155
      |||||::: |||
548 ...AsnHisThrCys...AspAsnAsnThrCys..... 556
      ::::: |||
1154 CGTTATGAATTTGGATGGTGGACCAACCGGTCAACACGTTCAATTGC 1105
      ||::: |||
557 .....AsnAsnHisThrLeuGlyAsnPro 564
      ::::: |||
1104 TATCATTTGTCACGCCAGTTTATCATTAATGGACATCCGAT..... 1065
      ::::: |||
565 HisPheTyrAsn.ProHisPheTyrAsnAsnThrLeuAsnMetProAsnA 581
      ::::: |||
1064 .....TCTGAACCGTTGATACCTTCGCCGCGTGTGCATTCCTGCTTT 1020
      ::||| |||
581 snLysLysGluThrHisAsnAsnPhe..... 589
      ::||| |||
1019 GTCGATGATGTTAAGCGTGTACTGTGGAA.....ATTCTAATGC 979
      ::::: |||
590 .....SerHisAsnAspThrGlnGluAsnAsnLeuMetLysAsnLys 603
      ::::: |||
978 TGATGATGTGCTCTTGATTAATATTGCTAATAATTGGAAATATGCCAA 929
      |||||::: |||
603 sAspGly.....LeuTyrLeuAsnThrLysSerTyrAspA 615
      ::||| |||
928 CAGATTTAATG...GCTGGCCAAGAGCTCCGATACCAATATGCGGAT 882
      ::::: |||
615 snAsnLeuPheGlyAla.SerAsnLysLeuThrSerHisSHISGLuAsnII 631
      ::::: |||
881 CGATCACAGCTTTCTATCATGCGCAGATCGATATACATTAAAGAAC 832
      ::::: |||
631 elysLysIleIleGluLeuAsnThrThrLysLeuValGluGluAlaGlyAsnA 648
      ::::: |||
831 AATAGCGGAATGTTGTCACCAACCAATGTTTCAGAACCCACAGAGATTGCGAG 782
      ::::: |||
648 sn.....SerLeuAsnAspIleAsnGluTyrAsnAsnAsnSerAsnAsp 662
      ::::: |||
781 CTGTTAAACAGAGTGGTGCCGACGACAAACCTGCTGCGACCTGCCGACATT 732
      ::||| |||
663 LeuAsnGluTyrPhe.....AspAsnLeuIleGluAsnAsnIleLe 676

```

```

731 CGTTACTCAGAAAAGATCTGCAGAACCGAGATA..... 695
   |||  ::  ::::::::::::::::::::|||
676 uSerTyrArgLysMetAsnIleLysAsnLysIleGlyThrLysPheI 693
694 .....TCATTGAGGTACGAATCATATCAACA..... 668
   ::|||::|
693 LeuMetAsnLysLeuMetTyrThrAsnValSerAsnAsnGluArgTyrArg 709
667 .....CCCTGAATTAAGCATGATATCAAGCTTTGCCA 633
   |||::|::|::|::|::|::|::|::|
710 TyrTyrLeuAspAspAsnLeuLysValSerTyrIleAsnGlnLeuArgLe 726
632 GTTGATT...TACGTACCGGTGCACCTTCGCAACATA... ATGGAACACC 589
   ::|||::|::|::|::|::|::|::|::|
726 uMetIleProTyrIleThrTyrCysLeuGlyLysIleAlaMetSerIleV 743
588 TGTAACTACTGTGCAGTACAAATGGAATCTGCATGTCACCATTTGGCT 539
   ::|::|::|::|::|::|::|::|::|
743 alphaTyrIlePheTyrIleLysPheAspIleSerTyrLeuLysLeuIle 759
538 TCTCATGTTTATGGGTTTAAAGCATTTGATTTGCTGCGCTCATTTATT 489
760 .....LeuTh 761
488 ACCATTTCGTTAAATTTCTGCCAATCAGAACGATAAATAATAGTGA 439
   ::|||::|::|::|::|::|::|::|::|
761 rAspTyrLysMetTyrPhe...LysLeuPheGlnHisLysAsnIleLeu 777
438 GAATCATTCAGAACAAATAAATACTGCCA.....TAT 407
   |||::|::|::|::|::|::|::|::|
777 helLeuValSerPheSerIleIleLeuGlnAsnThrIleIleSerPhePhe 793
406 AATTGCTTTCTTTATCATCTCTTAATACGTAATTTAGCTAAACAA 357
   |||::|::|::|::|::|::|::|::|
794 SerPheIle...PheLeuSerSerPhePheGlnValValLeuSerThrLeu 809
356 TATATAGTATGTAGG.....AAATTAATTAAGTGA 328
   ::|||::|
810 PheIlePheIleLysCysIleSerGlnPheLeuPheLeuLeuLeuVal 826
327 ATACAAATAGTGATAT.....TTTCATCAAAAC..... 300
   |||::|::|::|::|::|::|::|::|
826 lTyraSngluValPheGlnIlePheLeuArgAsnIleLysGlnProAspL 843
299 .....TTCTTCTATCGCTTTATAGCTTCGAAAAGCTTTATCA 261
   ::|||::|::|::|::|::|::|::|::|
843 ySerTyrAlaProTyrPhePheLeuThrPheAlaValIleProSerPheLys 859
260 TTAATCAGTAATCTTTATATGATCATCTATGTGAA..... 225
   ::|||::|::|::|::|::|::|::|::|
860 IleIleArgAsn...IleTyrPhePheLeuCysAlaLeuSerGlyArgGl 875
224 .....TGTTCATCATTAAGCCATGAATAGTTTGGTTGTTATTATCA 182
   ::|||::|::|::|::|::|::|::|::|
875 nPheIleAlaTyrIleIleArgProPheIleLysAsp..... 887
181 TCATTATCACTTGTCCTATTATTATCTCAAGATTATCATTTTGATATA 132
887 ..... 887
131 TATCAAAATTAATTAATCTTATGCCCATAATTTTANGGCATCATTTGCT 82
   |||  |||::|::|::|::|::|::|::|::|
888 .....LysAsnIleSerLysLeuProAsnPhePheAsnIleLysGluT 902
81 ATTCTGTAACAATTCGA 65
   |||::|::|::|::|::|::|::|::|
902 yTrAsnAsnAsnAsnAsn 907
seq_name: plf2:B26696
seq_documentation_block:
hypothetical protein 1 (Cyt-b-COII intergenic region) - Leishmania tarentolae mitochondrion
C:Species: mitochondrion Leishmania tarentolae

```

```

C:Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 24-Sep-1999
C:Accession: B26696
R:Simpson, L.; Neckelmann, N.; de la Cruz, V.F.; Simpson, A.M.; Faagin, J.E.; Jasmer,
J. Biol. Chem. 262, 6182-6196, 1987
A:Title: Comparison of the mexicircle (mitochondrial) genomes of Leishmania tarentola
A:Reference number: A92643; MUID:87194837
A:Accession: B26696
A:Molecule type: DNA
A:Residues: 1-443 <Sim>
A:Cross-references: GB:M10126; GB:J02707; GB:M10127; GB:M11022; GB:M64690; GB:N00030;
A:Note: the authors translated the codon ATT for residue 388 as Phe and TTC for resid
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC6
C:Superfamily: hypothetical protein 1 (Cyt-b-COII intergenic region)
C:Keywords: mitochondrion

alignment_scores:
Quality: 112.00 Length: 544
Ratio: 0.463 Gaps: 29
Percent Similarity: 44.485 Percent Identity: 19.853

alignment_block:
US-09-323-427-2/rev x B26696 ..
Align seg 1/1 to: B26696 from: 1 to: 443

1600 TTTATGTTCTGTACTACACTTATGCTATTTGATTTGATTCGCGTT 1551
   |||::|::|::|::|::|::|::|::|::|::|
3 LeuPheLeuTyrLeuIleHisIleIleLeuPheLeuLeuLeuTyrSerP 19
1550 GACAATGGTGTGAGGTGAGCCAG.....AAATGAATGTGGACCAAC 1507
   : ::|::|::|::|::|::|::|::|::|::|
19 eIleIleLeuCysAspPyrThrSerLeuPheTyrLeuSerPheAspLeuI 36
1506 TTCAAATPACATCAATTTTAATACAGTAATGATTCGTAAGACATGTTT 1457
   ::|::|::|::|::|::|::|::|::|::|
36 lSerPheLeuIleIleIleAsnIleIleLeuThrIleLeuAspSerTyr 52
1456 ATGTGAAGGTCTTTATGATCAGAAAGGTTGCCGTATATGATGAAGTGGA 1407
   ::|::|::|::|::|::|::|::|::|::|
53 lIecysPheIlePheLeuLeu..... 59
1406 CGTCAAGTTGCCGGAATTTCCATTCATTTGATTCGCAATGTCGCG 1357
   |||  |||  |||  ::|::|::|::|::|::|::|::|
60 .....LeuPheLeuPheCysPhePhePheLeuPheCysPheLeuAsn 74
1356 TACAGCATCTCTGAATCCAGCTGTATTTTGTGAACACACATGTTGTCA 1307
   ::|::|::|::|::|::|::|::|::|::|
74 heAspThrArgPheValPheMetIleIleMetGlnTyrIleIleIle 90
1306 TTTTCGTTTATCAATTAATTTGTTACCAAGTTGATGCGTGCATTCAGTA 1257
   |||  |||::|::|::|::|::|::|::|::|
91 PheMetPheLeuHis.....ValIleHisIleLeuP 101
1256 CAATGCTTTTACATGAAGCTGATAAACAGTTAGTCACAGATGAGGT 1207
   : ::|::|::|::|::|::|::|::|::|::|
101 eIleSerIleLeuPheGlnLeuPhe..... 109
1206 ATCTGAATACACACTGCTTTTCAAACTCAAAATG...TCCCGATCCAG 1160
   |||  |||::|::|::|::|::|::|::|::|
110 .....SerLeuLeuPheLeuIleLeuIleSerSerArgPheGly 123
1159 TATGCGTTATGAATTTTGGATGGTGGACCAACGCGTCAACAGTTCA 1110
   |||  ::|::|::|::|::|::|::|::|
124 TyrLysIleLeuIleLeuTyrP.....TyrTyrMetIleAs 136
1109 TTTGCTATCATTTGTCAGCAGCTTTATCAATAATGACATGCAATTCGA 1060
   |||::|::|::|::|::|::|::|::|::|::|
136 nLeuIleAsnPheIleLeuLeuPheValLeuLeuTyrTyrMetIleLeu. 152
1059 AACCGTTGATCTTCTGCGCGGTGTCCATTCTGCGTTGTC..... 1017

```

```

153 AsnTyrCysPheTyrLeu.....CysAspPhe.CysPheLeuIlePheA 167
1016 ..GATGATGTAACGCTGACTGTGGAATTCCTAATGCTGATGAGCT 969
167 spgIuGluTyrLeuGly.IleLeuCysLeuPheTyrIleLeuLeuIle 183
968 GCTCTGTAATATATTTGCTAATATTTGGAATATTCACAAACAGATTAAAT 919
183 uPheLeuLeuTyrIleAlaIleLeuIleLeuPheLeuGluGlnLeu.... 198
918 GGCTGGCCAAAGACTCAGCTATACAAATATGCGGATCGATCAGACTTT 869
199 .....TyrIleArgLeu..... 202
868 TCTATCATGCCAGATCAGTATTACCTTAAGAACCAATACGAGATG 819
203 .....GlyAla 204
818 GTTCGACCCACCAATGTTCAAGACCAAGATTCGAGCTGTTAAACAGG 769
204 lPhe..... 205
768 TGGTGCCGACAAACCTGCTGCGCTGCGCAACTTCGTTACTCAAGA 719
206 .....IlePheIleTyr..... 209
718 AAGATCTGCAGAACCGAATATCATGTATGTAACGATGATATCAAC 669
210 .....MetLeuThrPheTyrValIleuPheCys.. 218
668 ACCCTGAATTAGATGATATCAATCAAGCTTTCGACGATTGATTACGTCA 619
219 .....PheIleLeuIleIleLeuLeuIleCysPheIleTyrPheT 232
618 CCGTCACTTCTGCAACATATGACA.....ACCTGAATATAC 581
232 yrllellePheIleLeuLeuIle.IleIleGlnSerIleThrCysValle 248
580 TTGCTGCACTACAAATGATCTGCATCTCACATTGGCTTCTCAATG 531
248 uIleGlyLeuAsnSerPheAlaIleValIleSerLeuLeuPheValleSerV 265
530 TTATGCGTTTAAGCATTCATGATTGCTGCCGCTATTATTAACATTTC 481
265 aAlaAsnAsnPheSer.....PheLeuPheLeu 273
480 GTTAAATTTCTGCAAAATCAGAGCAATAAAATATGTTAGAAATCATC 431
274 IlePheIleSerThrIleAsnTyrIlePheTyrMetIleLeu..... 287
430 GAAGCAATAATAAAGTCCAT..ATATTTGGTTCTTCTTATATCATCC 384
288 .....AsnIleHisIleLeuIleTyrSerLeuSerLeuIleIleL 300
383 TTCTAATACTAATTTAGCTAACCAATATATAGTATAGGAATATAT 334
300 euIleIleLeuTyrTyrPhePheIleValIleTyrAsnMetPheAspIle..Ly 316
333 ACTGTAATACAAATAGATATTTTCATCAAAACTTCTCTATGCTTTT 284
316 s.....TyrAsnGlnAsnTyrPheLeuIleAsnPheIlePhe..... 328
283 ATAGCTTCTGAAGAGCTTATTCATTTATTCAGTAATCTTTATATCATAC 234
329 .....PheSerPhePheAsnPheAsnLeuIleSerIle 339
233 TATTTAATATGTTTCATCATTT.....AGCCATGATGATGTTTC 196
340 MetIleAlaCysIlePheLeuCysIleGlySerIleProIleValPheI 356
195 GTTT.....GTTATATCATCATTTATCAACTGTGCTAT..... 162

```

```

356 yPhePheLeuLeuValPheCysLeuLeuIleHisLeuSerTyrLeuGlyI 373
161 .....TTTATCTAACAGTTTATCATTTG 138
373 lEcysIleValPhePheSerIleIleTyrPheIleIleIleTyrIlePhe 389
137 TGATATATACAAATATACCTTATGTCGCAATTTTATATGCGCATCA 88
390 .....TyrPheArgLeuIleIleAsnIle..... 397
87 TTTCCTATCTCTGTAACAAATTC 66
398 .PheIlePheSerTyrGlnPhe 404
seq_name: p1r1.RNZQBF

```

```

seq_documentation_block:
DNA-directed RNA polymerase (EC 2.7.7.6) beta chain - Plasmodium falciparum plasmid
C:Species: Plasmodium falciparum
C:Date: 31-Dec-1990 #sequence_revision 15-May-1998 #text_change 11-Jun-1999
C:Accession: S72282; S10438
R:Wilson, R.J.M.; Denny, P.W.; Preiser, P.R.; Rangachari, K.; Roberts, K.; Roy, A.; W
J. Mol. Biol. 261, 155-172, 1996
A:Title: Complete gene map of the plasmid-like DNA of the malaria parasite Plasmodium
A:Reference number: S72277; MUID:96346169
A:Status: nucleic acid sequence not shown; translation not shown
A:Accession: S72282
A:Molecule type: DNA
A:Residues: 1-1024 <MIL>
A:Cross-references: EMBL:X52177; NID:99879; PIDN:CAA36427.1; PID:99880
A:Genetics:
A:Gene: rpoB
A:Note: this apparently degenerate plasmid is referred to as the apicoplast
C:Superfamily: DNA-directed RNA polymerase beta chain
C:Keywords: nucleotidyltransferase; plasmid; transcription

```

```

alignment_scores:
Quality: 112.00 Length: 665
Ratio: 0.373 Gaps: 28
Percent Similarity: 45.113 Percent Identity: 17.594

```

```

alignment_block:
US-09-323-427-2/rev x RNZQBF ..

```

```

Align seg 1/1 to: RNZQBF from: 1 to: 1024

```

```

1633 AATTACTAATCTTCTGAATATGATGCTTATGCTTATGCTTCTGACT 1584
70 AsnIleAsnAsnLeuLeuLysIleIleLeuThrLeuLysLeuAsnPheI 86
1583 ACACCTATTCGATGCTTATTCGATTCGCGGTGACATG..... 1544
86 eAsn...IleAsnLysIleIleLysPheAsnIleLeuIlePheIleLeuP 102
1544 ..... 1544
102 ropheIleTyrAsnAsnIleIleIleLeuAsnGlyLeuTyrLysThrCys 118
1543 .....GTTCGACAGCTGACCCAGAAATTTGAATGTCGACCAACTTCATA 1500
119 lIleGlnLeuPheLysLysAsnLysIlePheIleIleLysPheLysAs 135
1499 ACAATCA.....ATTTAATACAGCTAATGCAATTCGAGAGCATGT 1459

```

```

135 ..... ||| ..... |||
nasAsnLysAsnIleIleTyRValTyRIleTyRIleSerLeuGlyLeuA 152

1458 TTATGTGAAGGCTCTTATGATCAGAAGGTGGCGTAATGATGAGGTG 1409
    ||| ||| ||| |||
152 rGlie.....IlePheLysIleSerLys.....LeuAsnIle 162

1408 GACGTCAAGTGGCCGGAATTTCACTTCATTTGATTCAGCAATGTTGCG 1359
    ||| ||| ||| |||
163 AspCysTyRLeuAsnAsnPheLysPheAsnPheLeuIleLeuLeuLeuTy 179

1358 CGTACAGACGCTCGATCCACGCGATATTTTGTACACACACTGTGTT 1309
    .....
179 rLeuAsnAsn.....IleTyRIleAsnLysAsnIleSerL 191

1308 CATTGCTGTCATCATATTATTTGTACCAAGTTGATCGTCATATGAG 1259
    ||| ||| ||| |||
191 euPheIleTyRAsnAsnIleIleAsnLysIleLeuIle..... 204

1258 TACAAATGCTTTTACATGGAAGCGATATAAACAGTTAGTCACAGATTGAG 1209
    ||| ||| ||| |||
205 TyRAsnTyRIle.....LysPheIleTySerLysTyRAsnAsnIleAs 219

1208 GTATCTGAATCAACAAGTCTTTTCAACCAATTTGCCGATGCCAGT 1159
    : : : : ||| ||| ||| |||
219 nasnIleIleSerLeuLysLeuPheIleIleLysLeuAsnLysPheAsnA 236

1158 ATGCCGTTATGAATTTTGGATGGTGACCAACCGTCACAGTTCAAT 1109
    : : : : ||| ||| ||| |||
236 snIleTyR.....IleAsnLeuLeuAsn 243

1108 TTGCTA.....TCATTGGTCAGCCAGTTTATCAT 1080
    : : : : ||| ||| ||| |||
244 IleLeuPheSerIleLysLeuAsnPheSerTyRLeuAsnLysPheTyRI 260

1079 AATGGA.....CATGGATTCTGAACCGTTGATAC 1048
    ||| ||| ||| |||
260 eaAsnAsnIleTyRAsnLysLysPheTyRIleIleAsPAsnLeuLeu 277

1047 TTTCTCGCGGTTGTCATTCCTGCTTGTGATGATGATGACGATGATA 998
    : : : : ||| ||| ||| |||
277 leYserLys..... 280

997 CTGTGGAATTTCTAAATGCTGATGATGCTCTTGATTAATTTTGCTA 948
    ||| ||| ||| |||
281 .....LysTyRIleLysIlePheLysTyRIleLeuAsnIleAsnAr 295

947 AATAATTTGAATATCCACAGATTTAATGGCTGGCCAGAGACCTCAGT 898
    ||| ||| ||| |||
295 gaSnIleTyRAsn.....AsnIlePheLeuLeuAsnAsnLysLysT 310

897 ATACAATAATGCGGATCGATCACAGCTTTTCTATCAATGCCAGATCAGTA 848
    ||| ||| ||| |||
310 yRIleAsnIleIleLeuGlu.....AsnIleAsnIleAsnProuLeu 323

847 TTACACTTAAAGACCAATAGCGATGTGTTGACCACAAAT..... 806
    : : : : ||| ||| ||| |||
324 ValGlnTyRSerAspGlnValAsnAsnLeuSerGluIleAsnGlnLysP 340

805 .....GTTCAGAAC 796

340 euLysIleAsnMetIleThrThrGlyLeuAsnSerLysPheIleLeuAsnA 357

795 ACAAGGATTCGAGCTGTTAAACAGGTGGTGGCCGACGACAAAACCTGTG 746
    : : : : ||| ||| ||| |||
357 snAspLeuArgGlnLeuProArgAsnIleLeuGlyTyRIleSerLeuIle 373

745 CAG.....CTGCGCAACTGTGTT..... 728
    : : : : ||| ||| ||| |||
374 AsnThrAsnGlnGlyLeuThrCysGlyLeuValAsnTyRIleuThrThrAs 390

728 ..... 728

```

```

390 nIlePheLeuAsnLeuLysTyRIlePheValIleTyRTyLysHisIleP 407
727 .....TACTCAAGAAACATCTGCAGACCGGACAAATTCATTG 689
407 heTyRAsnArGtyRAsnPheLysLeuLeuAsnIlePheAsnLysAsn 423
688 ATGTACGAACTGATATCAACACCCCTGAAATTAGCAGTATATCAACGT 639
424 PheTyRAsnIleSerPheAsnAsnIleTyRIleuLysLysAsnIleAsnP 440
638 TTGCCAGTTGATTTAGCTACCGTGAC.....TTCTGCA 604
    : : : : ||| ||| ||| |||
440 eaSnLysThrThrIleuThrIleAsnLysAsnThrPheLysIleCysA 457
603 ACATAATGAGCAACCTGTAATFACCTGCTGAGTACAAATGGAATCTGCA 554
    ||| ||| ||| |||
457 snIleThrGlnAsnIleIleTyRIle.....Pro 466
553 TGTACCACTTGGCTCTCAATGTTTATGGGTTTAAACATTCATGATT 504
    : : : : ||| ||| ||| |||
467 PheAsnTyRIleuLeuSer.....ph 473
503 GCTGCCCTCATTTATACCATTTGCTTAAATTCGCCAATACAGAGCC 454
    : : : : ||| ||| ||| |||
473 eIleGlnAsnLeuIleProPhe...IleHisTyRAsnAspSerIleArgA 489
453 ATGAATAATGTTAGATCATGACGATTAATTAACCTGCCATATATA 404
    : : : : ||| ||| ||| |||
489 snLeuMetSerIleLysMetHisThrGlnIleValProIleIleTyPro 505
403 TTGCTTCTTCTTATCATCTCTTAATTAATTAATTTAGCTTAACATAT 354
    : : : : ||| ||| ||| |||
506 AsnLeuSerAsnIle.....IleThrAsnTy 514
353 ATAGTATGTGAATATATCTGTATACAAAT.....A 319
    : : : : ||| ||| ||| |||
514 rAsnPheIleLeuAsnLysTyRIleAsnHisLeuIleIleSerTyGlnG 531
318 GTGATATTTTCATCAAAACTCTTCTATGCTTTTATAGCTTGGAAG 259
    : : ||| ||| ||| |||
531 luGlyIleValIleTyRValSerCysIleLysIleIleIleArgasp... 546
268 CTATATTCATTATTCAGTAATCTTTATATGACATCATAT..... 230
    ||| ||| ||| |||
547 leuPheAsnArgGlnIleIleTyTyRIleuAsnAsnTyRIlysIleAs 563
229 .....G 229
563 nGlnAsnIleLeuLeuIleTyRIlysProIleValTyRValGlyLysV 580
228 TAAATGTTTCATCATTAGGCCATGAAATAGTTTGGTTGTTATATCATCA 179
    ||| ||| ||| |||
580 alAsnIleGlyGlnIleLeuAlaIleAsnSerAsnLeuLeuAsnSerGlu 596
178 TTATCAACTTGTCTATTTATTTCTAACAGTT..... 146
    ||| ||| ||| |||
597 TyRSerLeuGlyAsnAsnLeuLeuValGlyTyRIleSerTyRIleuGlyTy 613
145 .....ATCATTTGTGATATATACAAATATATACCT 115
613 rGluTyRIleAspAlaIleIleIleSerArgLysIleLeuTyRAsnAsn 630
114 TGTATTGCCCAATTTTATGGCATCATTTCCATATCTGTGAAC..... 71
    ||| ||| ||| |||
630 euTyRIleThrSerLeuHisLeuAsnIleTyGlnIleSerLeuAsnIleIle 646
70 AATTCACTT.....ATTGCATTTATGCAATTAATAAGTATTTTC 32
    ||| ||| ||| |||
647 AsnAsnIleProGlnIleCysSerIleAsnLeuSerLysMetTyR 661
seq_name: p1r2:S78177
seq_documentation_block:

```

hypothetical protein 717 - RecJinomonas americana (ATCC 50394) mitochondrion
 C:Species: mitochondrion RecJinomonas americana
 A:Variety: ATCC 50394
 C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 17-Mar-1999
 A:Accession: S78177
 R:Lang, B.F.; Burger, G.; O'Kelly, C.J.; Cedergren, R.; Golding, G.B.; Lemieux, C.; Sank
 Nature 387, 493-497, 1997
 A:Title: An ancestral mitochondrial DNA resembling a eubacterial genome in miniature.
 A:Reference number: S78127; MUID:97311393
 A:Accession: S78177
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-717 <LAN>
 A:Cross-references: EMBL:AF007261; NID:g2258325; PID:g2258376
 A:Experimental source: ATCC 50394
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1997
 C:Genetics:
 A:Genome: mitochondrion
 C:Keywords: mitochondrion

alignment_scores:
 Quality: 108.00 Length: 563
 Ratio: 0.414 Gaps: 28
 Percent Similarity: 46.359 Percent Identity: 19.538

alignment_block:
 US-09-323-427-2/rev x S78177 ..

Align seg 1/1 to: S78177 from: 1 to: 717

```

1471 TCGAGGAGCATGTTTATGTGAAGGCTTTATGATCAGAGAGGTTGCCGT 1422
      ::::::::::::::::::::
126 ThrLysAspLeuIlePheLeuLeuMetPheLeuAsnAspProIleLysI 142
1421 AATGATGAAGGTGAGCGTCAAGTGGCGAATTTCACTTCATTTGATTC 1372
      ::::::::::::::::::::
142 rAsnLysAsnLeuIleAsnAsnGlnAsnGlnLeuIleLysIlePheHis 156
1371 ATGCATGTGGCGGTACAGCATCTGAAATCCACGTGATTTTGTGA 1322
      ::::::::::::::::::::
157 .....TyrPheLeuPheLeuTyr 162
1321 CAACAACGTGTTCATCTTCCTTCATCATCATATTGTTACCAAGTTGAT 1272
      ::::::::::::::::::::
163 AsnAsnIleLysAsnThrAsnTyrVal.....LeuIle 173
1271 CGTGCAATGAGTACAAAGCTTTTACATGGAAGCTGATAAAGCTTAC 1222
      ::::::::::::::::::::
173 ethLysThrGlnLysAsnSerPheAlaIlePhe..LeuLeuLysGln...T 188
1221 TGGCAGATGAGGTATCTGAATACACAAGCTTTTCAAACTCAATG 1172
      ::::::::::::::::::::
188 hTrLysAsnLeuGlnTyrLleAsnIleGlnThrIleThrLysIleLysGln 204
1171 TCCCGATGCCAG.....TATGCCGTTAT 1149
      ::::::::::::::::::::
205 AsnTyrLeuGlnHisLeuPheAsnThrGlnAsnAsnLysTyrGlnIleAs 221
      ::::::::::::::::::::
1148 GAAATTTGGATGTGGACCAACCGTCAACAGTTTCAAT.....TATGCCGTTAT 1109
      ::::::::::::::::::::
221 nGlnSerPhePheTyrGlnGlnIlePhe.....PheAsnAsnTyrGlnIle 236
1108 .....TTGCTATCATTTGTACACCACTTATATCAATAATGACAT 1070
      ::::::::::::::::::::
236 rPlyLysGlnIlePhePheThrIleAspGlnIleLeuTyrLysThrHis 252
1069 GCGATTTGAAACGTTGATCTTTCTGCGCGGTGTCATTCCTCTCTT 1020
      ::::::::::::::::::::
253 LeuThrAsnLysLysGlnPhePheProAlaGlnIleAsnLeuThrAsn.. 268
1019 GTTCGATGATGTTAAGCTGATCTGTGAATCTAAATGCTGATGATG 970
      ::::::::::::::::::::

```

```

269 .....AspValGlnLeuPheLys..... 274
969 TGCCTTGATAATATTTGCTAAATATTTGGAATATTCACACAGATTAA 920
      ::::::::::::::::::::
275 .....ThrIleAsnAsnAsnThrIleTyrGlnIleLysIleProHis 289
919 TGGCTGGCCAGAGACCTCAGTATCAAAATATTCGCGATGACACAGCT 870
      ::::::::::::::::::::
290 GlnLeuAsnAsnLysIleLeuTyr...GlnLeuTyrIleGlnAsnGln 305
869 TTCTATCAATGCCAGATCAGTATTTCCATTAAAGAACCAATAGCGAATG 820
      ::::::::::::::::::::
305 nGlnIlePhe.....LysAsnLysLeuLeuTyrPlyAsnTyrIleProPhe 320
819 TGTTCGACCAACATGTTTCAGAACACACAGAGATTGCACTGTATAAACAG 770
      ::::::::::::::::::::
320 helleGlnLysGlnLysGlnLeuIleAspThrGlnIleIleAsnLys 336
769 GTGGTGGCCGACGAAACCTGCTGCAG.....CT 741
      ::::::::::::::::::::
337 TyrAsnLysLysIleAsnLeuLeuGlnGlnIleTyrTyrProLysVal 353
740 GCGCACTTCTGTTTACTCAGAAAGATCTG.....CAGAACCGAGAA 697
      ::::::::::::::::::::
353 IleuAsnTyrIleGlnTyrPheLysLysAsnLysIleAsnGlnAsnGlnLysI 370
696 TATCATGATGTCAGACCTGATATCAACACCCCTTGAATTAGCATGATA 647
      ::::::::::::::::::::
370 Le.....IleGlnTyrSerAsnLeuAsnIleGlnLeuAsnThrLeu 383
646 ATCAGCTTTGCCAGTTGATTTACCTCAC.....GTACACTTGGCA 603
      ::::::::::::::::::::
384 IleLysTyrIleArgLysThrAsnAspThrPhePheValIleIleLeuAs 400
602 CATTAATGCACACCTGTATATCTGCTGCGAGTACAAATGCAATCTGAT 553
      ::::::::::::::::::::
400 nSerLysIleSerLysLysTyrLeuLeu..... 409
552 GTACACATTTGGCTTCTCAATGTTATGGGTTTAAGCATTCATTCATG 503
      ::::::::::::::::::::
410 .....AsnIleLeuTyrLeu..... 414
502 CTGCGCTCATTTATACCATTTGCTTAA.....AATTCGTCCA 465
      ::::::::::::::::::::
415 .....GlnAsnLeuSerTyrPlyIlePheLeuGlnAspAsnPheLeuTh 429
464 AATCAGA.....AGCATAAATAATATGTAGATCAGTACGAAAGA. 425
      ::::::::::::::::::::
429 rIleGlnThrLeuIleLysAsnLysSerAsnIleIleThrAsnAlaG 446
425 ..... 425
446 IncysLeuAsnArgIleLeuAsnThrAsnThrAsnIleIleAsnLysAsn 462
      ::::::::::::::::::::
424 .....ATAATAAACCTGCCATATATATGCTTCTCTTATCATCC 384
      ::::::::::::::::::::
463 LysIleGlnIleLeuIleIleAspSerLeuPheThrAsnGlnGlnAsnAs 479
383 TTCTAATACATAATTTAGCTA..... 362
      ::::::::::::::::::::
479 rIleThrAsnLysPheSerLeuTyrIleThrAsnGlnThrLysGlnLysA 496
361 .....ACAATATATAGTAGTAGGAATAATATCTGTAATATACATAAGT 317
      ::::::::::::::::::::
496 snIleValAsnIleArgAsnGlnSerGlnLysLysSerAspAsnThrAsn 512
316 GATATTTTCATCAAAACTTCTCTATCGCTTTTATAGCTTGTGAAGAAGCT 267
      ::::::::::::::::::::
513 AspPheTyrGlnLeuLeuLysAsnThrGlnLysIleIleAsnLysGlnLys 529
266 TATTCATTTAT...TCAGTAATCTTT.....TATATGATACTATTG 229
      ::::::::::::::::::::
529 rIleTyrTyrLeuAsnIleValPheLeuThrGlnTyrValGlnArgIleA 546

```



```

228 TA...AATGTTTCATTCATGGCCATGAATAGTTCGTTGTTATATCA 182
      :::::::::::::::::::: ||||| |||
546 laaygaaspilleglnlysmetprohlscluphepelletryrphleuhis 562
      :::::::::::::::::::: ||||| |||
181 TCATTCATCAACTGTGCTATTATTCATACAGTTTATCATTTTGATAA 132
      :::::::::::::::::::: ||||| |||
563 tytleuyls..... 565
131 TAACCAAAATTTATACCTTGATTTGCCAATTTTATGCGATCATTTCC 82
      ||| ||| ||| ||| ||| ||| |||
566 .....Asnlystfyleuphecys...lletyrlsglnasnphrser 579
81 ATTCTGTA.....AACATTCACCTATTATTCGATTTGCAATTA 41
      :::::::::::::::::::: ||||| |||
579 yrlslylletyrlsasnasnlystfyleuphepelletryrphleuhis 595
40 AAGTATTTCATTGTCGAAAAA..... 2
      ::| ||| ||||| ||||| |||||
596 Gluilephe.....LysLysLysLysLysLys 605
seq_name: pir2:A40970

```

```

seq_documentation_block:
undulin 1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 13-Aug-1999
C:Accession: A40970
R:Just, M.; Herbst, H.; Hummel, M.; Duerkop, H.; Tripier, D.; Stein, H.; Schuppan, D.
A:Title: Undulin is a novel member of the fibronectin-tenascin family of extracellular m
A:Reference number: A40970; MUID:91373351
A:Accession: A40970
A:Molecule type: mRNA
A:Residues: 1-843 <JUS>
A:Cross-references: GB:W64108; NID:9340081; PIDN:AAA36794.1; PID:9340082
C:Superfamily: collagen alpha 1(XIV) chain; fibronectin type III repeat homology; von W
C:Keywords: glycoprotein
F:165-246/Domain: fibronectin type III repeat homology <FN3A>
F:255-338/Domain: fibronectin type III repeat homology <FN3B>
F:347-427/Domain: fibronectin type III repeat homology <FN3C>
F:436-520/Domain: fibronectin type III repeat homology <FN3D>
F:547-632/Domain: fibronectin type III repeat homology <FN3E>
F:641-723/Domain: fibronectin type III repeat homology <FN3F>
F:731-818/Domain: fibronectin type III repeat homology <FN3G>

```

alignment_scores:

Quality:	108.00	Length:	396
Ratio:	0.527	Gaps:	21
Percent Similarity:	51.768	Percent Identity:	21.212

alignment_block:

US-09-323-427-2/rev x A40970 ..

Align seg 1/1 to: A40970 from: 1 to: 843

```

1772 TTACCAAGTTTGAGGTGTCATATACACCGACTGCAGACACACACAAA 1723
      ||| :::::::::::::::::::: ||| ||| |||
228 leuthrglutyrglnllealvalphealalletrylalaahstphalase 244
      :::::::::::::::::::: ||| ||| |||
1722 CAACAACACACACACACACACACACACACACACACACACACACACAC 1673
      ::| ::| ||| ||| ||| ||| ||| |||
244 rgluglyleuthrglutyrglnllealvalphealalletrylalaahstphalase 261
1672 GAAGAAGACGAGGAAGCAATCTTACTTTTCTAATAAATGCAATTA 1623
      ::| ||| ||| ||| ||| ||| |||
261 sp.....LeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 269
1622 TCTTCGAATGATGATGATGATGATGATGATGATGATGATGATGATG 1573
      ::| ||| ||| ||| ||| ||| |||
270 Asnser...MetargvalystrpaspalavalproglalaSerglyty 285

```

```

1572 ATGTCCTTATTCGATTCGCGGTGACAAATGTCGCAAGT...GAGCCAG 1526
      ||| :::::::::::::::::::: ||| ||| |||
285 rleulleuleuylalproleuthrsluglyleualaglyaspdluylsg 302
      :::::::::::::::::::: ||| ||| |||
1525 AATGTAATGGGACCACTTCATTAACATCAATTTT..... 1488
      ||| :::::::::::::::::::: ||| ||| |||
302 lueelyslleglygluthrhisphrspllegluleuserglyleu 318
1487 ...AATACAGTAATGCATTCGAAGACATGTTATGTAAGTCTTTA 1441
      ||| ||| ||| ||| ||| |||
319 Proasphr.....GluTyrThrValThrValThrValThrValThr 331
1440 TGATCAAGAAGTGGCCGTAATGATGAGGTGAGCGTCAAGTTGCGGAA 1391
      :::::::::::::::::::: ||| ||| |||
331 egluglulalaseraspvalThrslugluluthrThrleualal 348
1390 TTTCACCTTCATTTGATTCATGCAATGTTGCGGTGACAGATCTGCAAT 1341
      ::| ||| ||| ||| ||| ||| |||
348 euserProPro...Argasnleuarglleaserasnvalglyserasnser 363
1340 CCAGTGTGATTTTGTACACACACACTGTTCATTTGCTTCATTCATTT 1291
      ||| ||| ||| ||| ||| |||
364 AlaArg.....LeuThrTrpAspPro... 370
1290 ATTGTTACCAAGTTCGATTCGATTCGATTCGATTCGATTCGATTCG 1241
      :::::::::::::::::::: ||| ||| |||
371 ...ThrserArgGlnlleasnlyltyrArgle.....ValTyrAsnA 384
1240 AAGCTATTAACAGTTAGTCACAGATGAGGTATCTGAATCACAACACT 1191
      ::| ||| ||| ||| ||| ||| |||
384 snAlaspllythrGluilleasnGluValGluValasprolethrThr 400
1190 GCTTTTCAACTCAATTCGCCGATG.....CCAGTATGCCGTTA 1150
      ::| ||| ||| ||| ||| ||| |||
401 PheProleuylglyleuthrProleuthrGluTyrThrIlealaleph 417
1149 TGAATTTTGGATGGTGGCAACCGGTCAACCACTT.....CAATTTG 1106
      ::| ||| ||| ||| ||| ||| |||
417 eserllytyrAspGluGlnser...GluProleuIleglyValalPhe 433
1105 CTAATCATTCGTCGACGATTTATCAATTAATGACATGCAATCTGAAC 1056
      ::| ||| ||| ||| ||| ||| |||
433 hrThrGluGluValProAlaGlnGlnTyrLeuGlnIleasplValalThr 449
1055 GTTGATCTTTCGCGCGGTTCCTTCCTTCCTTCCTTCGATGATGATG 1008
      ::| ||| ||| ||| ||| ||| |||
450 ThrAspserPheargValThrThrhisProleuserAlasplGlyle 466
1007 .....AAGGTGATACCTGTGAATTC 986
466 whlslyleuemetrrlleProvalTyrGlyLysThrsluglulaval 483
985 TAAATGCTGATGATGCTCTTGATTAATATTGCTTAATATTG... 939
      ::| ||| ||| ||| ||| ||| |||
483 alLeuLysGlu.....GluGlnaspserrhisvalIleglyLyu 497
938 .....GAATATCAACAGATTATTCGCT.....GGCCA 910
      ||| ||| ||| ||| ||| ||| |||
498 Progluthrsluglulavalserleuvalalaleuasrpsarglyse 514
909 AGAAGCTACGATATCA.....AATATGCGATGATCAGACGCTTT 869
      ::| ||| ||| ||| ||| ||| |||
514 rgluserclulvalValThrAlaValGlyThrThrleuasrserPheThr 530
868 TCTATCATGCCAGATGAGTATTACATTAACACCAATAGCAATGCT 819
      ::| ||| ||| ||| ||| ||| |||
531 ThrslugProAlaThrThrIlealPro...ThrThrserValThrserVa 546
818 GTTGACCAACATGTCGACACCAAGGATTCGAGCTGTTAAACAGG 769
      ||| ||| ||| ||| ||| ||| |||
546 lPheGlnThrcllyllearasn.....LeuV 555
768 TGGTGCCGACACAAACCTGCTGAGCTGCGC...AAGTTCGTTTACTCA 722

```

```

||||| ..... :||||| ..... |||
555 alvalglyaspdlthrthrserserleuavalylstrpaspleser 571
721 AGAAAGATCTGCAGAACCGGAATATCATGATG 686
572 AsperaspvalglnghphearvalthryMet 583
seq_name: plr2:C71607

seq_documentation_block:
hypothetical protein PFB0745w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 07-May-1999
C:Accession: C71607
R:Gardner, M.J.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
Pieria, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A>Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Release number: A71600; MUID:99021743
A:Accession: C71607
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1351 <GAR>
A:Cross-references: GB:AE001415; GB:AE001362; NID:g3845264; PID:g3845266; TIGR:PFB0745w
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0745w

alignment_scores:
Quality: 107.00 Length: 689
Ratio: 0.354 Gaps: 38
Percent Similarity: 43.832 Percent Identity: 19.303

alignment_block:
US-09-323-427-2/rev x C71607 ..
Align seg 1/1 to: C71607 from: 1 to: 1351

1755 GTCATTAACCCGACTGCAGCAACAACAACAACAACAACAACA 1706
||||| : : : : : ||||| : : : : : ||
323 lletyrlYsMetAsnArglYsAsnAsnNlleglNhIsAspAsp 339
1705 CAACAACAGCATTAAT.....AACCCATCAAGTGAGAGAGAAC 1662
||||| : : : : : ||||| : : : : : ||
339 nAsngluthrAsnAspVallylYsAsnNlletyrglYsLysA 356
1661 GAAGCAATCTAGATTCTTCTAAATCGAATTACTTAATCTTCTGA 1612
: : : : : : : : : : : : : : : : : : : : : : : : : : :
356 snlYsAsnNlletyrglYsAsnAsn..... 365
1611 GATGATTCGTCCTTATGCTTCTGTACTACACTTATGCAATGCTTAT 1562
365 ..... 365
1561 CGATTCGCGTTGACAAATGCTGCGAAGTGACCGAATGATGGA 1512
366 .....A 366
1511 CCAACTTCAATACATCAATTTTAATACAGTAAGCATTCGAAGACA 1462
: : : : : ||||| : : : : : : : : : : : : : : : : : : : :
366 snAsnAsnAsnAsn. AsnlyshIsMetAsnlysserlIeserThrAs 382
1461 TGTG.....TATGGAAGGT.....CTTATGATCAAGAGGTT 1427
: : : : : ||||| : : : : : : : : : : : : : : : : : : : :
382 nlleuAsnlystYrllleYsAsnGlnhIsllleValThrlyshValI 399
1426 GCCGAATGATGAAGGTGACGTCAAGTTGCCGAATTCACCTTCATT 1377
||||| : : : : : : : : : : : : : : : : : : : :
399 leArgThrAspdlulYsLysglulLeupne..... 409
1376 GATTCATCAATGTTCCGCGTACAGATCTCTGATCCACGCGATTTT 1327
||||| : : : : : : : : : : : : : : : : : : : :

```

```

410 ...PheCysThrPheValAsnMetThrLeuLeuTyrglullel 425
1326 TGTAAACAACACTGTTGCA.....TTTCGTTTC 1298
||| : : : |||
425 euphetyrlYsAsnIleserThrAsnNlleYsllleAsnTyrglul 441
1297 ATC.....CATTATTTGTACCAAGTGTGATGCATATGCA 1260
||| : : : : : : : : : : : : : : : : : : : :
442 lleAspAspThrTrpAsnAsnIlelleThrAsnIlelleTyrlle 458
1259 GTACATGCTTTTACATGGAAGCTGATAAACAATTAGTCACAGATTGA 1210
458 s...AsnAsnIlePromeGlnArglIleYsLysglulThrhis..... 471
1209 GGTATCTGAATCAACAAGCTCTTTCAAACTCAATTTGCCGATGCCAC 1160
||||| : : : : : : : : : : : : : : : : : : : :
472 ....LeuGlnserIlelleSerleuTyrseryleuthrValleuAsn 486
1159 TATGCCGTTATGAAA.....TTTGATGCTGAGCAACCGGTCAA 1119
||||| : : : : : : : : : : : : : : : : : : : :
487 TyrSerlyslLeuTyrglulAsnIlePheTyrllePheGlnArgserValAs 503
1118 CCAAGTTCATTTCTATCATTTGTCAGCCAGTTTATCATTAATG...A 1073
||||| : : : : : |||||
503 lIlellehIsAspLeuphelyshIsAsnMetArglYsllleAsnIleMet 520
1072 CATGCGATTCGAAACCGTTGATACTTCTGCGGGCTTGCCATTCCTGC 1023
||| : : : : : |||
520 hrPheAspGlnleuLysAsnAspLeuAsnValSerPheValAsnMetCys 536
1022 .....TTTGATGATGATGTAACGGTGATCTGT 994
||||| : : : : : |||||
537 AsnAspAspAsnAsnAsnAsnAsnAspAspAsnAsnAsnAsn 553
993 GGAATTCCTAATGCTGATGATGCTCTTGATTAATATTGGTAATA 944
||| : : : : : |||
553 pAsnAsnAsnAsnAspAsnAsnGlyAspAsp.....AsnA 566
943 ATTGGAAATATCAACAGATTAAATGCTGCGCAAGAGACTCATATAC 894
||| : : : : : |||
566 sn.....AspAspAsnValIle 572
893 AATATGCGGATCGATCAACAGCTTTCTATCAATCCAGATCATATTAC 844
||||| : : : : : |||||
573 lYsTyrlYs.....HisSer...AsnValGlnProlyslYsAsnL 586
843 CATTAAGAACCAATAGCGAATGCTGTGACCACAAATGTCGAACAC 794
: : : : : : : : : : : : : : : : : : : : : : : : : : :
586 yVallyslYsAsnMetTyrlAsnThrPhehIsArgAsnIleYsPheLys 602
793 AAGGATTCGAGCTGTTAAACAGGTGCGCCAGCAAAACCTGCTGCA 744
: : : : : : : : : : : : : : : : : : : : : : : : : : :
603 TyrlYsGlnAsnIleValhIsAsnTyrlLeuAsnlyslleAspProleu 619
743 GCTGCGCACTTC.....GTTTACTCAAGA 718
619 uTyrlAsnAsnPheleupheValTyrlValProAspLeuLeuTyrlSer 636
717 AAGATCTGCAGAACCGGAGATATCATTTGATGTAACGATATACACA 668
636 spAsnCyslThrAsp..... 640
667 CCCTTGAAATATAGGAGATTAATACAGCTTTGCCAGTTGATTACGTCAC 618
: : : : : : : : : : : : : : : : : : : : : : : : : : :
641 .....MetPheThrLeuAspdlulLeuThrlyslLeuLeuTyrlale 654
617 CGTGCATCTTCGAACATATGACAAACCTGTAATCTTCTGCTGACAGTACA 568
||||| : : : : : |||
654 uSerTyrlTyrl.....Glnlysg 660
567 AAATGAATCT.....GCATGCACCATTTGGCT 539
: : : : : : : : : : : : : : : : : : : : : : : : : : :
660 lIleGlnlysglnlyslYsAsnAsnlyslArglYsllleTyrlhIsllleYs 676

```

```

538 TCTCAATGTTATGGGTTAAGCATTCATGTCGCGTCATATT 489
    ::::: |||||
677 ASP.....LlelleIleSerLeu 683
488 ACCATTTCGTTAAATTCGTCGAATCAGAGG..... 455
    |||::: ::||| :::::
683 uProtYr...ValasnThrIleValGluArgLnlPhePheLysLeu 699
454 ..CATAAAATATGTTAGAAATCGAAGCAAAATAAACTGCCATAT 407
    ::|||::|::|
699 alaAsnLysAsnAsn.....AsnIleCysSerLysIleLysAsn 711
406 ATATTCGCTTCTCTTATCATCTTCTTAATACATAATTTAGTACAAA 357
    ||| ::::: |||||
712 IleGluThrCysAsnLeuAsnIleTyrAsnAsn.....ValAspPr 725
356 TATATAGTAT.....GTAGCA.....AATATT 334
    ::|::|::|
725 oValValTyrLysAsnLysLeuAlaValGlyLysMetClnLysAsnAsn 742
333 AC.....TGTAAATACATAAGT..... 317
    |||::: ::|||
742 YrAspLysAsnThrCysSerIleLeuSerSerTyrLysAsnTyrLeuAsn 758
316 .....CATATTTCATCAAAACTTCTTATCGCTTTTATAGC 279
    ::::: |||||::|
759 IleCysAsnAspAsnThrTyrValAlaIleHisSerSerIle...TyrCysI 774
278 TTCTGAAGAAGCTTATTCATATTCAGTAATCTTTTATGCAATCTAT 229
    ::|||::|::|
774 eGluLysAsnLeuSerHisLeuLeuAsnIleTyrTyrGlnHisLysL 791
228 TAAATGTTTCATCATTAGGCCATGAATAGTTTGGTTGTTATATGATCA 179
    ||:::|::|::|
791 alaSprLeuYmet.....PheTyrIleLeuThrPheLeu 802
178 TATTCACACTGTCCTATTATTCATACAGTTTATGATTTGATATAT 129
    |||::: ::|||
803 LeuAlaMetProLysLysLysTyrIleAspLeuIleIlePheSerAsn 819
128 CACAAT.....TATACCTGTATATGCCCA 103
    |::|
819 eIleAsnAlaLeuSerLysMetCysTyrThrTyrGluMetTyrValAl 836
102 TTTTATGGGCATCATTTCTATCTGTAACAACATCATTATTTGCATT 53
    ::|||
836 euphe.....TyrPheValAsnLys.....ValCysGly 845
52 ATTGCAATTAATAAGTATTTCATTTGT..... 26
    |||
846 IleArgIleSerGluTyrValLeuSerLysTyrPhePheArgAsnGly 862
25 .....GAAAAA 13
862 uValLeuYstrHrValGluGluGluLysGluGluGluGluGly 879
12 AAAAAA 2
    ::|||::|
879 IuLysGluLys 882
seq_name: p1r2:G71613
seq_documentation_block:
hypothetical protein PFB0470w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 07-May-1999
C:Accession: G71613
R:Gardner, M.J.; Tetteijn, H.; Garrucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600, MIMD:99021743

```

```

A:Accession: G71613
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1346 <GAR>
A:Cross-references: GB:AE001397; GB:AE001362; NID:93845192; PID:93845193; TIGR:PFB047
A:Experimental source: clone 307
A:Gene: PFB0470w

alignment_scores:
  Quality: 105.50      Length: 719
  Ratio: 0.319         Gaps: 48
  Percent Similarity: 46.036      Percent Identity: 21.280

alignment_block:
  US-09-323-427-2/rev x G71613

Align seg 1/1 to: G71613 from: 1 to: 1346

1710 AACACAAACACAGCAATTAATACCCCATCAAGTGAGAGAAAGACAG 1661
    |||::|::|::|::|::|::|::|
649 AsnHisGluAsnSerArgAsnLysIleAlaAsnThrPrlAsnLysIleG 665
1660 AAGCAATCTAGTTTCTTAATAATGCA..... 1633
    |||
665 uHisAsnAsnLeuAspAsnLysAsnAsnAsnMetAspPheAsnAsnMet 682
1632 .....ATTACTAAATCTCTGAATGATGAT 1606
    |||
682 eThrSerProLeuTyrTyrTyrTyrTyrTyrTyrTyrAspAsnAsp 698
1605 TCG...TCTTATTCGCTTCTGTAC.....TACACTTATGCAATTG 1568
    |||::|::|::|::|
699 SerLysGlyTyrCysAsnLeuPheGluIleLeuTyrGlyTyr..... 712
1567 CTATATCGATTCCGGTTGACAAATGCTGCAAGGTGACGCCAATTTGAA 1518
    ::|||::|
713 .....AsnSerGlyTyrAsn..... 718
1517 TGTGACAACACTCAATTAATCAATTTTATACAGTAAATGATTCGA 1468
    ::|||::|::|::|::|::|::|
718 eUtrThrSerPheSer.SerLeuSerTyr.Val...ValLysTyrAsnG 733
1467 AGCAGATGTTTATGTAAGGCTTTATGATCAAGAAGTTGCCGTAATG 1418
    ::|||::|::|::|
733 IuGlnMetPheLeu...LysLysPheLysAspSerLysInsEr..... 746
1417 ATGAAAGTGGACGTCAAGTTCGG..GAATTCACCTTCATTTGATTCA 1371
    ::|||
747 .....GluValPrlAsnAsnPheGluIleHisLeuAspAs 758
1370 TGCAAATGTTGCGGTACAGCATCTCGAATCCACGCGATTTTGTATAC 1321
    ::|
758 n.....IleSerAspLysIleLeuLysIlele.....G 768
1320 AACAACTGTTGTCATTTGCTTATTCATCATTTATTTGTTACCAAGTTGAT 1271
    ::|||::|::|::|::|::|
768 IuGlnAsnLeuAsnHisGluAsnMetLysTyrIleIleHisAsnLeuMet 784
1270 GTG.....CATATCGAGTACAAATGCTTTTAC 1245
    ::|
785 IleSerLeuSerLeuCysAspIleLeuLysTyrLeuAsnLeuTyrAlaLeuCy 801
1244 ATGGAAGCTGATAAACAGTTAGTCACAGATTGAGG.....T 1207
    ::|
801 sphePheIleLeuLysGluAsnTyrTyrTyrLeuSerIleAspAsnLeuT 818
1206 ATCTGAATATCAACAAGCTTTTCAAACTCAATGTCGCCGATGCCAG... 1160
    |||||
818 YrLeuTyrLeuGluIleLeuHisArgMetLysIleTyrAsnHisAspIle 834
1159 ...TATGCGGTTATGAATAATT.....TGAGTGTGGA 1131

```

```

835 PheYrSerIleMetGluTyrIleAsnThrHisValHisAlaLeuGluSe 851
      |||.....|
1130 CCAACCGGTCAACAGTCATTTGCTTTCAT..... 1098
      |||.....|
851 rGlnLysMetLysIlePheLeuLeuSerTyrAsnIlePheGlnLysM 868
      |||.....|
1097 ..GGTCAGCCAGTTATCATTAATGACATCGCATTCGAAACCGTTGAT 1050
      |||.....|
868 eLaspAsnProValaspMetLysGluMetCys.....Asp 879
      |||.....|
1049 ACTTCTGCGCGTTCCTTCATTCCTGCTTTCATGATGATGATGATGATG 1000
      |||.....|
880 PhePheLeuSerSerAsnAsnLysIleGluLysGluAsnGlyAsnAsp 896
      |||.....|
999 TACTGTGAAATTCATAATGCTGATGATGCTGCTGATTAATATTG. 951
      |||.....|
896 P.....LeuMetLeuGlyLysCysThrHisGluLysAsnLeuT 909
      |||.....|
950 .....CTAATAATTTGGAA 936
      |||.....|
909 rPlysLeuProThrAspIleGluIleLysGlnAsnLeuIleAsnLeuGlu 925
      |||.....|
935 ...TATCCACAGATTAAATGCTGCG.....CAAGAGC 904
      |||.....|
926 AsnPheGlnLysGluLeuLeuSerAsnAsnAspAsnAspLysMetGluPh 942
      |||.....|
903 TCACGTATACAAATATGCGATC.....GATCACAGCTTTTCTATC 863
      |||.....|
942 eHis.AspsAsnAsnCysAsnIleIleGlyHisAspLysPhePheSerAsn 958
      |||.....|
862 AATGCCAGATCAGTATTACCA.....TTAAGAACCA 831
      |||.....|
959 AsnAspGluAsnLysIleLysLysGluLysTyrPheAsnLeuLysAsnGlu 975
      |||.....|
830 AATA.....G 826
      |||.....|
975 uLleMetValPheLysLysIleGluLysThrGluThrLeuProCysThrL 992
      |||.....|
825 CGAATGCTGTCAGCAC.....AATGTCAGAACCAACAG 791
      |||.....|
992 euAsnIleTyrAspTyrIleAsnPheLeuLeuIleLeuIlePheTyrGln 1008
      |||.....|
790 GATTCGAGACTGTTAAACAGAGTGCGCCAGCAAAACCTGCTGCAGCT 741
      |||.....|
1009 CysAsnAsnLysIleLysGluCysAspGluLysIleAsnLeu..... 1022
      |||.....|
740 GCGCACTTCGTTTACTCAGAAAGAT.....CTGCAGA 706
      |||.....|
1023 ...AsnPheLeuPheSerLysAspGluAsnValIleIleThrIleGln 1038
      |||.....|
705 ACCGGAGAAATATCATTCATGTACGAA.....CTGATATCAACACCC 665
      |||.....|
1038 snGlu.....MetTyrGluLysAsnAsnLysIleLysAsnPro 1050
      |||.....|
664 TTGAAATTAGC.....ATGATTAATCAAGCTTTGCCAGTT 630
      |||.....|
1051 CysLysTyrValLysAsnLysGlnTyrMetLeuAspLysTyrSerGluMe 1067
      |||.....|
629 GATTACGTCACCGTCACCTTCGCAACATAATGACACACCTGTAATACT 580
      |||.....|
1067 t.....LeuLysGluAsnLeu..... 1072
      |||.....|
579 TGTCTGACTACAAATGGAATCTGCATGTCACCATTTGGCTTTCATGAT 530
      |||.....|
1073 .....PheAsnIleGluSerSerLeuIleGlnLeuPheSer..... 1084
      |||.....|
529 TTATGGGTTTAAGCATTCGAT.....TGATGCTGCC 498
      |||.....|
1085 IlePheValAsnLeuLeuGluLysGlyGluAspAspLysGluLeuPheVal 1101
      |||.....|
497 GTCATTATTACCATTTGTTTAATTTGTCCAAAATCAGAGCATTAATA 448
      |||.....|

```

```

1101 lAsnGlnIleMetPheIleLeuAspPheIleLysIleIleAsnGluLysV 1118
      |||.....|
447 ATAATGTTAGATCATCGAACATA..... 422
      |||.....|
1118 aTyrIleAsnIleMetLysIleValLysLysMetLysAsnTyrAspGlu 1134
      |||.....|
421 ..ATAAAGTCCATATATATGCTTCTTCCTTATCATCTCTTATATA 375
      |||.....|
1135 AsnIleLysArgLysAsnTyrPheThr.....TyrSerLysAsnLys 1149
      |||.....|
374 CTAATTTAGCTTAACAAATATATAGTAT.....GTAAG 343
      |||.....|
1149 sTyrPheGlnLeuLysLysIleAspLeuGluTyrIleAsnSerAsnIleA 1166
      |||.....|
342 GAATATATCTGTAATACATATAGTATTTTCATCA..... 303
      |||.....|
1166 snAsnLysLysLysAsnThrTyrAsnAspPhePheAsnGluAsnAsn 1182
      |||.....|
302 ...AAGTTCTTCATGCTTTTATAGCTTCTGAAAGCTTATTCATTAT 256
      |||.....|
1183 lLeAsnTyrArgTyrGlnTyrGlnSerValHisLysAlaIleGlnLeuPh 1199
      |||.....|
255 CAGTAAT.....CTTTATATGCACTATTTGTAATGT..... 222
      |||.....|
1199 eSerAspAsnIleIleArgTyrSerHisAsnGluLysIleAsnThrHisT 1216
      |||.....|
221 .....TTCATCATTAGGCCA..... 207
      |||.....|
1216 yrlsAsnAsnLysTyrIleIleLysAspIleLysThrPheTyrLysLeu 1232
      |||.....|
206 .....TGAAATAGTTCGTTTATATATATCATTCAT..... 177
      |||.....|
1233 AspsAsnPheLeuIleSerAspIleLeuLeuIleLeuGluLysGlnAsnLys 1249
      |||.....|
176 .ATCAACTTGTCTATTTATTTCTA.....A 152
      |||.....|
1249 sGluGlnIlePheTyrPheLeuLeuPheTyrProPheGluLeuLysGlnT 1266
      |||.....|
151 CAGTTATCATTTTGATATATATATACAAATTAACCTGTATTTGCCCAT 102
      |||.....|
1266 hValIleHisIle.....LysAsnAsnThrPheLeuPheAsn 1278
      |||.....|
101 TTT...TATGGCATCATTTTCTTATTCGTATAAACAATTCATTATTGCA 55
      |||.....|
1279 TyrLysTyrAspGluThrPheLeuPheAsnMetGluIleLeuPheLeuTyr 1295
      |||.....|
54 T 54
1295 r 1295

```

seq_name: plr2:S55098

```

seq_documentation_block:
probable membrane protein YMR216c - yeast (Saccharomyces cerevisiae)
N/Alternate names: hypothetical protein YMR261.10c
C/Species: Saccharomyces cerevisiae
C/Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 06-Feb-1998
C/Accession: S55098
R/Dedman, K.; Brown, D.; Bowman, S.
submitted to the EMBL Data Library, June 1995
A/Reference number: S55089
A/Accession: S55098
A/Molecule type: DNA
A/Residues: 1-742 <DED>
A/Cross-references: EMBL:249809; NID:9854459; PID:9854468; MIPS:YMR216c
A/Experimental source: strain AB972
C/Genetics:
A/Gene: SGD:SKY1
A/Cross-references: SGD:S0004829; MIPS:YMR216c
A/Map position: 13R
C/Keywords: transmembrane protein
F:241-257/Domain: transmembrane #status predicted <TM>

```



```

1094 CAGCCAGTTTATC.....ATPAATGACATGCGATTCGAAACCGT 1054
||||:|||||
1205 userSerPheIleasnTyrgInIleLysThrAsnSpmEtLeuTyraSnL 1222
||||:|||||
1053 TGATCTTTTCGCGGGTTCATTCCT...GCTTGTGATGATGGTA 1007
||||:|||||
1222 euLeuLeuSgluLysSerAlaTyrgInAsnInIleuSgluLysasnTy 1238
||||:|||||
1006 ACGGTGATCTGGAATTCGAATGCTGATG....GATGGCTCTT 963
||||:|||||
1239 IleuIleIleuTyrgValLeuLeuSerMetTyrgLysLysMetAl 1255
||||:|||||
962 GATAATATTTGCTAAATAATTGATATCCACAGATTTAATGGCTGG 913
||||:|||||
1255 alIeTyrgPheArgSerPheValTyrgAsnInIleLysValSerLysL 1272
||||:|||||
912 CCAGAGAGCTCAGCATACAAATATGCGATGCATCAC..... 875
||||:|||||
1272 ysaSnAlaPheAlaTyrgThrLeuThrArgValAsnSerIleuValLeu 1288
||||:|||||
875 ..... 875
1289 TyrgLuarArgIleLysSerPheIlePheSerLysLeuLysPheasnTy 1305
||||:|||||
874 .....ACGTTTCTATCAATGCCAGAT 853
||||:|||||
1305 rAspAsnValSerTyrgPheCysPheThrMetTyrgLleTyrgLeuArg 1322
||||:|||||
852 CAGTATTA.....CCATTAAAGAACCAATATACG 824
||||:|||||
1322 rgIleLeuPheGlyTyrgLeuArgIleArgAspAsnArgIleAsnIle 1338
||||:|||||
823 AATGTGTTCGACCACATGTT..... 803
||||:|||||
1339 AsnValIleGluLysAsnValTyrgArgLeuValLysLeuIleSerLys 1355
||||:|||||
802 ....CAGAACCCAGAGATTCGAGCTGTTAAACAGGTGGTCCGACAC 757
||||:|||||
1355 eSerAspAsnHisLysTyrgAsnAlaPheLeuLys..... 1366
||||:|||||
756 AAACCTGCTGCAGCTGCGCAACTGTTTACTCAAGAAAGATCTGCAG 707
||||:|||||
1367 .....LeuGlnLysTyrgValTyrgLugInAsnGluLysLys 1378
||||:|||||
706 AACCGAGAGATA.....TCATTGATGTAC.....GAACGTATAC 672
||||:|||||
1379 AsnLysMetIleCysAspAsnLeuIleTyrgAlaAsnAsnGluLeuCy 1395
||||:|||||
671 AACACCTTGAATTAAGCATG.....A 649
||||:|||||
1395 nasLLeuAspLysIleAlaIleGluLysGlyIleAsnGlnIleAspCys 1412
||||:|||||
648 TAATCAAGCTT..... 638
||||:|||||
1412 euIleLysPheLysArgLysGluCysLeuMetLysTyrgPheTyrgThr 1428
||||:|||||
637 .....TGCCAGTTGATTACGTCACCGTCACCTTTCG 606
||||:|||||
1429 LysGlyProGlnIleAsnThrGluArgPheTyrgTyrgCysIleArgTy 1445
||||:|||||
605 CAACATAATGCACAACTGTATACTGCTGCAGTACAAATGGAATGTG 556
||||:|||||
1445 sSerIlePheSer.....PheValLeuAsnLysIleIleGlnLysL 1459
||||:|||||
555 CATGTCACCATTTGGCTTCTCAATGTTATGGGTTAAGCATTCGATGCA 506
||||:|||||
1459 ysaValGlnHisIlePhePheGlnPhe..... 1467
||||:|||||
505 TTGCTGCCCTCATTTATTACATTTTCGTTAAAT..... 473
||||:|||||
1468 ValLeuLysThrLeuGlnArgAsnAsnLysAsnArgLeuThrHisAla 1484
||||:|||||

```

```

472 .....TTGCTCCAAATCAGAAAGCATPAAATAATGTTA 439
||||:|||||
1484 elysLeuLeuGlnValIleValGlnLysLysGluLysSerValIleA 1501
||||:|||||
438 GATATATGAGCATATATAAAGCTCCATATATATTC..... 401
||||:|||||
1501 spValLeuGlnLeuTyrgAspLysTyrgProTyrgIlePheGlnTyrg 1517
||||:|||||
400 .....GTTTCTTTATCATGCTTCTTAATTAATTTTACTTAAC 360
||||:|||||
1518 LeuThrLysIleGluValPheValIleCysValGlnAsnPheVal... 1532
||||:|||||
359 AATATATATGATGTAGAAATATTAATTAATTAATTAATTAATTAAT 310
||||:|||||
1533 .....ThrLeuTyrg 1535
||||:|||||
309 TCATCAAACTCTCTCATGCTTTTA 283
||||:|||||
1536 AsnArgLysLeuLeuLeuAsnPheLeu 1544
||||:|||||
seq_name: p1r2:E71604
seq_documentation block:
Hypothetical protein PFB0870w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 07-May-1999
C:Accession: E71604
R:Gardner, M.J.; Tetteh, H.; Garucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.; Perera, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MID:99021743
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2380 <GAR>
A:Cross-references: GB:AE001421; GB:AE001362; MID:93845293; PID:93845297; TIGR:PFB087
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0870w
alignment_scores:
Quality: 104.50 Length: 528
Ratio: 0.415 Gaps: 24
Percent Similarity: 47.727 Percent Identity: 19.886
alignment_block:
US-09-323-427-2/rev x E71604 ..
Align seg 1/1 to: E71604 from: 1 to: 2380
1749 AACACCGACTGCAGCAACACAAACAAACAAACAAACAAACAAACAA 1700
||||:|||||
490 AsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsn 506
||||:|||||
1699 CAGCAATATAAACCCCATCAAGTGAGG...AAGAAGACGAGAGCAATC 1653
||||:|||||
506 nasAsnAsnAsnGlnTyrgLysLysIleLysAsnLysAsnLysAsnL 523
||||:|||||
1652 TTAGTTTTCATAAAATCGAATTTACTTAATCTCTGAAATGATGATTG 1603
||||:|||||
523 ysaSnLysLysLysAsnLysLeu.....Asn 531
||||:|||||
1602 TCTTATTCCTTCTGTACTACACTTATTCATTTGCTTAT...TCGATTC 1556
||||:|||||
532 AsnTyrgAsnAspAsnPheValSerValAsnGlySerTyrgAspAsnTy 548
||||:|||||
1555 CGTTGACAATGGTGTGAGAGTGAGCCAGAAATGAAATGTGACCAACT 1506
||||:|||||
548 erIleAspAsnVal..... 553
||||:|||||
1505 TCATATACATCATTTTAATACAGTAATGATTCGAAAGACATGTTTA 1456
||||:|||||

```

```

554 .....lleasnaspGluIleargGluLysLysasnLysG1 567
1455 TGCGAAGGCTTTATGATCAGACAGTTGCCGTAATGAGACGTGCAC 1406
567 uValLysIleMetValaspLysasnAspThrGluLysaspGlysnL 584
1405 GTCGAGTTGCCGGAATTCACCTTCATTGAT .....TCATGCAAT 1365
584 LysLys .....TyrAspThrSerTyrSerPheasn 593
1364 GTTCCGCGTACACGATCT .....CTGATCCACG 1336
594 IleLysasnThrLysSerLysValPheTyrLysasnTyrValLysaGly 610
1335 TGGTATTTTGTACACACAGCTGTTCATTTGCTTTCATTCATTATTTG 1286
610 sGlyMet.IleLysGlnGlnHisasnLleHisasnThrHisasnIle 626
1285 TTACCAAGTTGATCGTATCGATGACATGCTTTTACATGGAAGCT 1236
627 HisasnThrHisasnMetHisasnThrHisasnIleHisasnGluLysVa 643
1235 GATPAAACAGTATAGTCACAGATTTGAGTATCGAATCACAACCTGCTT 1186
643 Val .....LeuLeuaspThrLysLysValaspPheMetasnL 658
1185 TCAACTCAATTTGCCGATGCCAGTATGCCCTTATGAAATTTGGATG 1136
658 euGlyLleSerPheSerProIaGlyLeuLeuIle ..... 669
1135 GTGACCAACCGGTACACAGTTCAATTTGCTATCATTTGGTCAGCCAGT 1086
670 .....ProTyrHisLeuGlyValSerSerLeuLeuIleGluLysas 683
1085 TATCATAAATGACATGCGATTCGTAACCGTTGATACCTTTCGCCGCT 1036
683 lleLeuasnMetHisThrSerIleIaGly .....SerSerAlaG 697
1035 TGTTCATCTCTGCTTTGTGATGATGTAACCGTGATCTACTGTGGAATTC 986
697 LysIleLysIaLysCysLysSerValGlyLeuSerValasnLysCys 713
985 TAAATGCTGATGATGCTCTTGTATTAATAT .....TGCTA 948
714 TyrPheLeuIleGluasnIleIleSerasnValTyrLysHisGlyCys.. 729
947 AATAATTGGAAATATCAACAGATTTAATGCTGGCCAGAAGCTCAGCT 898
730 ...TyrGlnLysLeuGluasnIle ..... 736
897 ATACAATATWTGCGATCGATCGATCTTTTCTATCATGCCAGATCAGTA 848
737 .....LeuasnIleGluLeuasnLysTyrLeuTyrGluaspSerTyr 750
847 TTACCATTAAGAACCAATAGCGATGTGTGACACACAAATGTTACAGA 798
751 IleTyrLeuasnAsnArgIleGlyasnValPheValGly ..... 763
797 CCACAAGATTCGGAGCTGTAAACAGGTGTGCCGACGACAAAACCTGC 748
764 .....IleThrGlnIleLeuProTyrTyrLysLysL 774
747 TGCACGTGCGCACTTGTCTACTCAAGAAAAGATCTGCAGAACGGGAGA 698
774 euasnIleasnAsnPhe...TyrAspAspAsnAspLeu ..... 785
697 ATATCATTTGATGTACAGTATCAACACCTTGAATTTAGCATGAT 648
786 IleSerAlaIleIleIleIleSerCysasnIleProMetTyrLeuSer..... 800
647 AATCAGCTTTGCCAGTGTATTACGTCACCGTGCATCTTGCACACATTA 598

```

```

801 .....SerAsnIlePheValAsnPheArgAsnLysLysCysI 813
597 TGCAACACCTGTAAATCTGCTGCAGTACAAATAGATGCATGTCAC 548
813 leaspGly .....PhePheSerThrLysLysLysaspPheGlyCysPro 827
547 CATTTGGCTTCTCAATGTTTATGGGTTAAGCATTTGATGATTCGCC 498
828 AsnThrArgThrGlu .....ArgIleI 835
497 GTCATTTATACCATTTGCTTTAAATTTGCTTCACAAATAGACGATAAA 448
835 eLysValSerProPheaspSerAspTyrValGlyIle...GlyasnLysA 851
447 ATATGTTAAGATCATGACGACATTAATAA .....CTGCCATAT 407
851 snAsnSerValIleSerProHisLeuIleLysTyrAsnHisIleLeuPhe 867
406 ATATTGCTTTCTTCTATATCATCTTCAATTAATTTAGCTAACAA 357
868 LeuPheIleCys ..... 871
356 TATATGATATAGAAATATATCTGATATACATTAAGATATTTTCA 307
872 .....ValLysasnIlePheHisLysTyrIleasnLleuTrpI 885
306 TCAAACTCTCTTATCGCTTTTATAGCTTCTGAAAAGCTTATTCATTAT 257
885 leGluLysaspTyrLeuPheLeuIleGluasnLeuLysaspIleGlu 901
256 TCAGTATCTTTTATATAGCATTAATTTGTA 227
902 ArgLysIlePheaspTyrTyrThrPheVal 911

seq_name: p1r2:S57242

seq_documentation_block:
twichin - Caenorhabditis elegans
N:Alternate names: myosin-regulating protein
N:Contains: protein kinase (EC 2.7.1.-)
C:Species: Caenorhabditis elegans
C:date: 28-Oct-1995 #sequence_revision 24-Oct-1997 #text_change 18-Jun-1999
R:Benian, G.M.; Kiff, J.E.; Neckelmann, N.; Moerman, D.G.; Waterston, R.H.
submitted to the EMBL Data Library, February 1993
A:Description: Additional sequence complexity within twitching of Caenorhabditis eleg.
A:Reference number: S57242
A:Accession: S57242
A:Molecule type: DNA
A:Residues: 1-6835 <BEN1>
A:Cross-references: EMBL:LI0351
A:Experimental source: var. Bristol
R:Benian, G.
submitted to the EMBL Data Library, November 1989
A:Reference number: S07571
A:Accession: S07571
A:Molecule type: DNA
A:Residues: 792-6839 <BEN2>
A:Cross-references: EMBL:X15423; NID:96897; PIDN:CAA33463.1; PID:96898
A:Experimental source: var. Bristol
R:Benian, G.M.; Kiff, J.E.; Neckelmann, N.; Moerman, D.G.; Waterston, R.H.
Nature 342, 45-50, 1989
A:Title: Sequence of an unusually large protein implicated in regulation of myosin ac
A:Reference number: S06797; MUID:90044042
A:Accession: S06797
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 806-1175;1178-1998,'Y',2000-3040,'I',3042-3335,'I',3337-5693;5696-6359,'I
A:Cross-references: EMBL:X15423
A:Experimental source: var. Bristol
R:Benian, G.M.; L'Hernault, S.W.; Morris, M.E.
Genetics 134, 1097-1104, 1993
A:Title: Additional sequence complexity in the muscle gene, unc-22, and its encoded p

```

A: Reference number: S57218; MUID: 93387664
 A: Accession: S57218
 A: Molecule type: DNA
 A: Residues: 2-99,108-194, 'Q', 196-206,374-468,658-753 <BEN4>
 A: Experimental source: var. Bristol
 C: Comment: Lack of unc-22 leads to a constant twitching of the body muscles.
 C: Genetics:
 A: Gene: unc-22
 A: Map position: IV
 A: Introns: 18/3; 69/3; 143/2; 176/3; 264/2; 387/3; 413/2; 471/1; 516/3; 550/3; 582/3; 60152/3; 6691/3; 6776/1; 6808/3
 C: Superfamily: twitchin; fibroectin type III repeat homology; immunoglobulin homology;
 C: Keywords: ATP; autophosphorylation; duplication; muscle; phosphotransferase; serine/th
 F: 806-898,899-990,993-1083,1084-1175,1178-1273,1474-1567,1770-1864,2066-2158,2358-2450,2
 F: 56-5790,6263-6356,6386-6478,6541-6635,6649-6742,6745-6838/Region: motif 2
 F: 1274-1372,1373-1473,1568-1670,1671-1769,1865-1964,1965-2065,2159-2258,2259-2357,2451-2
 F: 4215-4313,4314-4416,4416-4516,4612-4710,4711-4811,4908-5009,5010-5109,5110-5210,5399-
 F: 5940-6197/Domain: protein kinase homology <KIN>
 F: 5948-5956/Region: protein kinase ATP-binding motif
 F: 5971/Active site: lys #status predicted

alignment_scores:
 Quality: 104.00 Length: 536
 Ratio: 0.479 Gaps: 26
 Percent Similarity: 40.485 Percent Identity: 19.403

alignment_block:
 US-09-323-427-2/rev x S57242 ..

Align seg 1/1 to: S57242 from: 1 to: 6839

```

1749 AACACCGACTGACGACACACACAAACAAACAAACAAACAAACAA 1700
      |||||
3290 AsnThrHisIleAspAsnAsn.ThrAspAsnAsnThrIleAspIleThr 3306
      |||||
1699 CAGCA.....ATATAACCCCA 1683
      |||||
3306 ysserThrAlaArgAlaAspserGlyLysTyrLysIleValAlaThrAsn 3322
      |||||
1682 TCACCTGGAGGAGAGACAGACAGACATCTTACGTTTCTTAATATC... 1635
      |||||
3323 GlusergGlyLysAspGluHisGluValAspValAsnIleLeuAspIle 3339
      |||||
1634 .....GAATTTACTAATCTTCTG 1616
      |||||
3339 oGlyAlaProGluGlyProLeuArgHisLysAspIleThrLysGluSer. 3355
      |||||
1615 AATATGATGATTCGTTTATGTTCTGTACTACTTATTCGATGTGCT 1566
      |||||
3356 .....ValValLeuLys 3359
      |||||
1565 TATTCGATTCGCGTTCGACAAATGCT.....GT 1540
      |||||
3360 TrpAspGluProLeuAspAspGlyLysProIleThrAsnTyrValVa 3376
      |||||
1539 CGAAGTGACCCAGAA.....ATTGAATGTGACCAACTTCAA 1502
      |||||
3376 IGLuLysGluGluAspGlyLysGlyTyrValProCysGlyLysIleSer 3393
      |||||
1501 TAAACAATCAATTAATACAGATGATTCGATTCGACGACATGTTAT... 1455
      |||||
3393 sPrThr...SerLeuLysValAsnLysLysSerGluGlyHisGluTyrLys 3408
      |||||
1454 .....GTGAAGGTCTTTATGATCAACAAGTTCGCGTAAATGATGAAG 1411
      |||||
3409 PheArgValLysAlaVal..... 3414
      |||||
1410 TGGAGCTGAAGTTCGCGAATTCACATTCATTCATTCATTCATTCAT 1361
      |||||
3415 AsnArgGluGlyThrSerAlaProLeuThrSerAspHisAlaIleVal 3431
      |||||
1360 CGCGTACACGATCTGTGATTCACGATGATGATTTTGTAAACAACACT 1311

```

```

3431 IalysAsnProPheAspGluProAspAla..... 3440
      |||||
1310 GTCAATTCGTTTCATCCATTAATTTGTTACCAAACTGAT..... 1272
      |||||
3441 .....ProThrAspValThrProValAspTrpAspLysAs 3452
      |||||
1272 ..... 1272
      |||||
3452 PHisValAspLeuGluThrLysProProAlaAsnAspGlyAlaProI 3469
      |||||
1271 ..CGGCAATGACGATGACATGCTTTTACATGCAAGCTGATTAACACT 1224
      |||||
3469 IeAspAlaTyrIleValGluLysLysAspLysPheGlyAspTrpValGlu 3485
      |||||
1223 AGTGCACAGATTCAGGTATCTGAATTCACAACTGCTTTCAACTCAAT 1174
      |||||
3486 CysAlaArgValAsp...GlyLysThrThrLysAlaThrAlaAspAsn 3501
      |||||
1173 TGTCCGATGCCAGTATGCCGTTATGAAT..... 1143
      |||||
3501 uThrProGlyGluThrTyrGluPheArgValLysAlaValAsnLysAla 3518
      |||||
1142 .....TTGATGTGACCAACCGGTCAACCACT..... 1113
      |||||
3518 LyrProGlyLysProSerAspProThrGlyAsnValAlaValAlaLys 3534
      |||||
1113 ..... 1113
      |||||
3535 ArgMetAlaProLysLeuAsnLeuAlaGlyLeuLeuAspLeuArgIle 3551
      |||||
1112 .....CAATTCGATCATGTCGACCACT..... 1086
      |||||
3551 salagIlyThrProIleLysLeuAspIleAlaPheGluGlyLysProAla 3568
      |||||
1085 ..TATCATTAATGACATGCGATTCGTAACCGTTGATCT..... 1047
      |||||
3568 rovalAlaLysTrpLysAlaAsnAspAlaThrIleAspThrGlyAlaArg 3584
      |||||
1046 .....TTCGCGC 1039
      |||||
3585 AlaAspValThrAsnThrProThrSerSerAlaIleHisIlePheSer 3601
      |||||
1038 GGT.....GTCCATTCCTGCTTGTGCGATGATGTTACG 1004
      |||||
3601 aValArgGlyAspThrGlyValTyrLysIleIleValGluAsnGluHis 3618
      |||||
1003 GT...GATCT..... 996
      |||||
3618 LylLysAspThrAlaGluCysAsnValThrValLeuAspValProGlyThr 3634
      |||||
995 .....GTGCAATTCGTAATGCTGATGATGCTGCT 964
      |||||
3635 ProGluGlyProLeuLysIleAspGluIleHisLysGluLysThrIle 3651
      |||||
963 TGAATTAATTTGCTAATTAATTTGCAATATCCACAGATTAATGCGT 914
      |||||
3651 u.....AsnTrpLysProProThrAspAsnGlyLys 3662
      |||||
913 GCSAAGAAGCTAC...GTATACAATATGCGATGATGACAG..... 873
      |||||
3662 hrAspValLeuHisTyrIleValGluLysMetAspThrSerArgGlyThr 3678
      |||||
872 .....CTTTGATATCAATGCCAGATCGATATTAACATTA 838
      |||||
3679 TrpGluGluValGlyThrPheProAspLysThrAlaLysValAsnLys 3695
      |||||
837 AGAACCAATAGCCGA...TGGTTGACACCAATGTTGACACCAAG 791
      |||||
3695 uValProGlyLysGlyTyrAlaPheArgValLysAlaValAsnLeuGlu 3712
      |||||
790 GATTGAGAGCTGTTAAACAGGTGTCGCGACAAACCTGCTGCACT 741
      |||||

```


OM of: US-09-323-427-2 to: SwissProt_38.* out_format : pfs

Date: Apr 16, 2000 4:59 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+.n2p.model -DEV=xlp
-O=/cgn2.1/USPROT.spool/US09323427/runat_14042000_170514_19931/app-query.fasta.1
-DB=SwissProt_38 -QFMT=fastan -SUFFIX=backtrans_rsp
-GAPOP=12.000 -GAPEXT=4.000 -MIMATCH=0.100 -LOOPECL=0.000
-LOOPEXT=0.000 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-YGAEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blomsum2 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
-THR_SCORE=pcct -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
-MINLEN=0 -MAXLEN=1000000 -USER=US09323427 -NCPU=6 -ICPU=3
-NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-323-427-2

Query length: 1779

Database: SwissProt_38.*

Database sequences: 83229

Database length: 29864866

Search time (sec): 138.700000

Score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Document
SwissProt_38:CTUT1_CAEEL	1420.50	2216.62	1.6e-116	424	1 003755 caenorhabditis elegans
SwissProt_38:ROB_PLAFA	112.00	152.35	0.0652	1024	1 P21421 plasmodium falciparum
SwissProt_38:YOH3_CAEEL	109.50	155.32	0.0921	485	1 009276 caenorhabditis elegans
SwissProt_38:PSL1_BOVIN	106.50	153.40	0.1581	369	1 P21809 bos taurus (bovine)
SwissProt_38:RK65_YEAST	104.50	143.62	0.2755	742	1 003656 saccharomyces cerevisiae
SwissProt_38:AMW_PIG	103.50	137.89	0.3708	1150	1 P12021 sus scrofa (pig)
SwissProt_38:NM5M_APLII	103.00	144.04	0.3498	554	1 P34855 apis mellifera (honey bee)
SwissProt_38:TF2_BACSU	102.50	140.82	0.4091	716	1 P17889 bacillus subtilis
SwissProt_38:MUC5_BOVIN	100.50	139.96	0.5809	563	1 P98081 bos taurus (bovine)
SwissProt_38:AMC4_DICDI	100.00	144.63	0.5669	317	1 P14198 dictyostelium discoideum
SwissProt_38:YAG3_HVSA	99.50	136.09	0.7450	717	1 P16492 herpesvirus saimiri
SwissProt_38:YAF3_SCHPO	97.50	133.90	1.157	649	1 010211 schistosomus haematophili
SwissProt_38:YCF2_MARPO	97.00	121.80	1.57	2136	1 P09975 marchantia polymorpha
SwissProt_38:CALU_PSEAE	96.00	139.55	1.23	279	1 059633 pseudomonas aeruginosa
SwissProt_38:CYAA_DICDI	96.00	124.19	1.76	1407	1 003100 dictyostelium discoideum
SwissProt_38:PGS1_MOUSE	95.50	136.11	1.45	369	1 P28653 mus musculus (mouse)
SwissProt_38:ATC_PLARK	95.50	124.70	1.89	369	1 P47853 rattus norvegicus (rat)
SwissProt_38:PGS1_HUMAN	94.50	134.57	1.77	368	1 008853 plasmodium falciparum
SwissProt_38:YVIG_ECOLI	94.50	132.85	1.85	441	1 P21810 homo sapiens (human)
SwissProt_38:AMD2_XENLA	94.50	126.34	2.14	875	1 P12890 xenopus laevis (afriid)
SwissProt_38:SRB8_YEAST	94.50	122.70	2.38	1427	1 P23648 saccharomyces cerevisiae
SwissProt_38:YV9P_YEAST	94.00	122.87	2.52	1161	1 P47179 saccharomyces cerevisiae
SwissProt_38:TLK3_CAEEL	94.00	118.22	2.80	1895	1 P41951 caenorhabditis elegans
SwissProt_38:PM55_NEUCR	93.50	131.61	2.24	426	1 P23351 neurospora crassa
SwissProt_38:PYRI_SCHPO	92.00	113.47	4.35	2244	1 009794 schistosomus haematophili
SwissProt_38:DYHG_CHIRE	91.50	106.11	5.60	4485	1 039575 schistosomus haematophili
SwissProt_38:GND_MOUSE	91.00	114.75	4.99	1663	1 P01027 mus musculus (mouse)
SwissProt_38:YV61_MERJA	90.50	124.82	5.46	2515	1 P25823 dirosophila melanogaster
SwissProt_38:NM5M_NEUCR	90.50	121.98	4.59	553	1 058956 methanococcus jannaschii
SwissProt_38:YV15_YEAST	90.50	114.58	4.44	715	1 P05510 neurospora crassa
SwissProt_38:G156_PAPPR	90.50	109.31	6.14	1559	1 P14772 saccharomyces cerevisiae
SwissProt_38:PEP3_YEAST	90.00	118.82	5.36	2715	1 P1837 parametium primaurel
SwissProt_38:NRK1_YEAST	90.00	117.28	5.55	918	1 P27801 saccharomyces cerevisiae
SwissProt_38:STU1_YEAST	89.00	121.61	5.98	1080	1 P38692 saccharomyces cerevisiae
SwissProt_38:VE1_HPV6	89.00	121.61	5.46	1513	1 P38198 human papillomavirus
SwissProt_38:IRA2_YEAST	89.00	105.76	8.54	630	1 080957 human papillomavirus
SwissProt_38:PGS1_CANFA	88.00	125.12	5.95	3079	1 P19158 saccharomyces cerevisiae
SwissProt_38:SSR3_RAT	88.00	124.49	6.55	369	1 002678 canis familiaris (dog)
SwissProt_38:YB17_ANCFU	88.00	119.42	7.36	363	1 P30938 rattus norvegicus (rat)
				619	1 029441 archaeoglobus fulgidus

SwissProt_38:SYA_MYCPN - 88.00 115.86 7.99 900 1 P75368 mycoplasma pneumoniae
SwissProt_38:RMR_CANCA - 87.50 124.35 7.14 339 1 P21358 candida glabrata
SwissProt_38:UL52_HSV7J - 87.00 114.71 9.68 861 1 P52468 herpes simplex vir
SwissProt_38:AMD_HUMAN - 87.00 113.54 9.94 974 1 P19021 homo sapiens (human)

seq_name: SwissProt_38:CTUT1_CAEEL

seq_documentation_block:

ID	CTUT1_CAEEL	STANDARD:	PRT:	424 AA.
AC	003755; Q18693;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	CUTICLIN 1 PRECURSOR.			
GN	CUT-1 OR C47G2.1.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;			
OC	Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2.			
RX	MEDLINE: 91323673.			
RA	SEBASTIANO M., LASSANDRO F., BAZZICALUPO P.;			
RT	"cut-1 a Caenorhabditis elegans gene coding for a dauer-specific			
RT	noncollagenous component of the cuticle."			
RL	Dev. Biol. 146:519-530(1991).			
RM	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2.			
RA	PALMER S.;			
RL	Submitted.			
CC	-1- FUNCTION: COMPONENT OF THE CUTICLES. IT FORMS A RIBBON			
CC	APPROXIMATELY 2 MICRONS WIDE RUNNING ALONG THE LATERAL LINES			
CC	UNDERNEATH THE ALAE. CONTRIBUTES TO THE FORMATION OF EXTRACELLULAR			
CC	ENVELOPES PROTECTING THE ORGANISM FROM THE ENVIRONMENT.			
CC	-1- SUBCELLULAR LOCATION: SECRETED.			
CC	-1- TISSUE SPECIFICITY: LOCALIZED IN A NARROW BAND NOT EXPOSED			
CC	TO THE EXTERIOR, RUNNING UNDERNEATH THE ALAE.			
CC	-1- DEVELOPMENTAL STAGE: EXPRESSED DURING DAUER LARVA FORMATION.			
CC	-1- DOMAIN: THE SMALL REPEATS A-A-P-[AVI] ARE ALSO PRESENT IN MANY			
CC	PROTEINS CONSTITUTING THE PROTECTIVE ENVELOPE OF OTHER SPECIES.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M55997; AAA27995.1; ALT_INIT.			
DR	EMBL; Z49125; CAA88934.1; -			
DR	PIR; A49772; A49772.			
DR	HSSP; P04002; 1ATF.			
DR	WORMPEP; C47G2.1; CE02165.			
KW	Cuticle; Signal; Repeat.			
FT	CHAIN	1	18	POTENTIAL.
FT	DOMAIN	19	424	CUTICLIN 1.
FT	REPEAT	302	323	4 X 4 AA REPEAT OF A-A-P-[AVI].
FT	REPEAT	302	305	1.
FT	REPEAT	307	311	2.
FT	REPEAT	312	315	3.
FT	REPEAT	320	323	4.
FT	CONFICT	27	27	G -> P (IN REF. 1).
FT	CONFICT	29	29	P -> G (IN REF. 1).
FT	CONFICT	306	306	A -> G (IN REF. 1).
FT	CONFICT	309	309	MISSING (IN REF. 1).
FT	CONFICT	389	390	CL -> SS (IN REF. 1).
FT	CONFICT	395	396	FA -> LP (IN REF. 1).
FT	CONFICT	401	402	IG -> MR (IN REF. 1).
FT	CONFICT	401	402	
FT	SEQUENCE	424 AA;	45168 MM;	08E05529 CRC32;

```

792 AGAATTGGACCTGTAAACAGCTGCT..... 765
793 ||||| ||||| ||||| ..... 768
794 nglpneqlvalavallysglnalaglydylalanlslalaia 291
795 .....GCCGA 759
796 ..... 764
797 laalaproglinalglyvalgluvalglinalalaprovalaia 291
798 GCAAAACTGCTGCAGCT..... 741
799 ||| ||||| ||||| ..... 744
800 AlaAlaProValAlaAlaProValAlaAlaAlaAlaProAlaIa 324
801 ..... 740
802 .....GCCCACTTCGTTTACTC...AGAAGAGCTC. 711
803 AlAAlaProValAlaAlaProValAlaAlaAlaAlaAlaProAlaIa 341
804 ..... 710
805 ..GCAGACC GGAAATATCATTGATGTACAGAACCGATATCACACCTT 663
806 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| 341
807 helqylgnslngslnglylleuAspValArgValGlutleasnThr 357
808 GAATATGACGATGATAATCAAGCTTGCCAGTTGATTAATTCACCTGC 613
809 ..... 358
810 Asplemetglugly.....AlaserProSerAlaProgluaAlaAla 372
811 ||||| ||||| ||||| ||||| ||||| ||||| ||||| 372
812 ACTTCGTCACATTAATGACACCTGTAACTTCTGTCGACATACAANA 563
813 ||||| ||||| ||||| ||||| ||||| ||||| ||||| 372
814 alevalserglu....GluserValArgArgAlaThrSerThg 387
815 ..... 562
816 GAATTCGATGTCACCATTTGGCTTCATGTTTAAAGGTTAAACATT 513
817 ||||| ||||| ||||| ||||| ||||| ||||| ||||| 387
818 lyllecysLeuthrProillelyphehaseerphenueuglyilegylthr 403
819 ..... 512
820 GCATTGATTCGTCGCGTCATTAATCAATTCGTTTAAATTCGTC 465
821 ..... 404
822 llevalatrhAlaleuSerAlathrllepherYValAlaArgPro 419
823 .....
seq_name: SwissProt_38::RPOB_PLAFA
ID RPOB_PLAFA STANDARD; PRT: 1024 AA.
seq_documentation_block:
AC P21421;
DT 01-MAY-1991 (Rel. 18, Created)
DI 01-OCT-1996 (Rel. 34, Last sequence update)
DI 01-NOV-1997 (Rel. 35, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE BETA CHAIN (EC 2.7.7.6).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BW(C10);
RA MEDLINE: 95107345.
RA GARDNER M.J., GOLDMAN N., BARNETT P., MOORE P.W., RANGACHARI K.,
RT STRATH M., WHITE A., WILLIAMSON D.H., WILSON R.J.M.:
RT "Phylogenetic analysis of the rpoB gene from the plastid-like DNA of
RT Plasmodium falciparum."
RL Mol. Biochem. Parasitol. 66:221-231(1994).
RN [2]
RP SEQUENCE OF 328-1024 FROM N.A.
RX MEDLINE: 91187055.
RA GARDNER M.J., WILLIAMSON D.H., WILSON R.J.M.:
RT "A circular DNA in malaria parasites encodes an RNA polymerase like
RT that of prokaryotes and chloroplasts."
RL Mol. Biochem. Parasitol. 44:115-124(1991).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC RNA(N).
CC -1- MISCELLANEOUS: THIS RNA POLYMERASE IS ENCODED ON A CIRCULAR DNA.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.

```



```

489 snleuMetSerIleIysMetHisThrGlnIleValProIleIleIleTyrPro 505
403 TTCCTTCTCTTATCATCTTCTAATACCTAATTTAGCTACACAAATAT 354
506 AsnLeuSerAsnIle.....IleThrAsnTyr 514
353 ATGATATAGTAATATATCTGTAATACATA.....A 319
514 rAsnphelIleuAsnLysTyrLeuAsnHisIleIleIleSerTyrGlnG 531
318 GTCGATATTTTTCACAAACTTCTATCGCTTTTATAGCTTCGAAAG 269
531 lGlyIleValIleTyrValSerCysIleLysIleIleIleArgSP... 546
268 CTTATTCATTTATTCAGTAATCTTTATATAGCATACTAT... 230
547 LeupheAsnArgIlnIleIleTyrTyrLeuAsnAsnTyrLysLysIleAs 563
229 .....G 229
563 nGlnAsnIleLeuIleTyrLysProIleValTyrValGlyLysIle 580
228 TAAATGTTTCATCATAGGCCATGAAATAGTTGTTTATATATCATCA 179
580 AlsnIleGlyGlnIleLeuAlaIleAsnSerAsnLeuAsnSerGlu 596
178 TTAATCACTTGCTCCATTTATCTACAGTT..... 146
597 TyrSerLeuGlnAsnAsnLeuValGlyTyrGlySerTyrLeuGlyTyr 613
145 .....ATCATTTGTGATTAATATGACAAATATATACCT 115
613 rGlyTyrGlnSPAlaIleIleIleSerArgLysIleLeuTyrAsnIle 630
114 TGTATGCGCCATTTTATGCGATCATTTCTTATCTGTAAC..... 71
630 euryrThrSerLeuHisLeuAsnIleTyrGluIleSerLeuAsnIleIle 646
70 AATTCACCTT.....ATTGCAATATTCGAATTAATAAGTATTTTC 32
647 AsnAsnIleProGlnIleIleCysSerIleAsnLeuSerLysMetTyr 661

seq_name: SwissProt_38:YQH3_CAEBL

seq_documentation_block:
ID YQH3_CAEBL STANDARD; PRT; 495 AA.
AC 009276;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEETICAL 55.1 KD PROTEIN C43C3.3 IN CHROMOSOME X.
GN C43C3.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilitidae;
Rhabdilitinae; Rhabdilitoidea; Rhabdilitidae; Peloderinae; Caenorhabditis.
RN Rhabdilita; Rhabdilitoidea; Rhabdilitidae; Peloderinae; Caenorhabditis.
RC STRAIN-BRISTOL N2.
RP SEQUENCE FROM N.A.
RA SUSTON J.;
RT Submitted (DEC-1994) to the EMBL/GenBank/DBI databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 247067; CAAB7330.1;
DR MORPEP; C43C3.3; CE01525.
KM Hypothetical protein; Transmembrane; ATP-binding.
FT TRANSMEM 11 31 POTENTIAL.

```

```

FT TRANSMEM 426 446 POTENTIAL.
FT NP_BIND 171 178 ATP (POTENTIAL).
SQ SEQUENCE 495 AA; 55068 MW; 09c77468 CRC32;

```

```

alignment_scores:
  Quality: 109.50      Length: 415
  Ratio: 0.559        Gaps: 24
  Percent similarity: 47.229      Percent identity: 21.446

```

alignment_block:

US-09-323-427-2/rev x YQH3_CAEBL ..

Align seg 1/1 to: YQH3_CAEBL from: 1 to: 495

```

1745 CCGACATCGGCAACAACAACAACAACAACAACAACAACAACAACAAGC 1696
    |||
26 ProIleValMetPhePheAsnAsnGlnAlaAsnThrSerLeuIleHt 42
1695 AATTAATACCCCATCAAGTGA.....GGAGACAGACAG 1661
    |||
42 sPheLeuGlyProHisSerProSerSerLeuPheValAlaHisThrA 59
1660 AACCAATCTTATGTTTCTAAATATGAA.....TTTACTAATCT 1620
    |||
59 sPheLeuValAlaTyrAlaThrSerGlnAlaGlyAlaPheValArgGlu 75
1619 TCTGAATGATGATTCGCTTATGCTTCTGTAATCACTTATGCTAT 1570
    |||
76 MetArgAlaLysLeuGluLeuVal.....AspCysIleAlaAs 88
1569 GTCCTTATTCGATTCGCGTTGACAAATGCTGACAGGTGACAGAAATTT 1520
    |||
88 pSerPheThrValValLeuAsn.....LysSerAspProIleValM 102
1519 AA.....TGTGACCACTTCATTAATACATCAATTTTAATACAGT 1479
    |||
102 eLargMetIleSerAsnProLysSerGlnProVal..... 113
1478 AATGCATTCGAAGACATGTTTATGTAAGGCTTTTATGATCAAGAG 1429
    |||
114 .....ValTyrValTyrGlnHisLysThrArgHisPr 124
1428 TTGC.....CGTAATGATGANGGTGACGTCACAGTGCAGGAATTT 1388
    |||
124 cCysGlyThrSerMetLysAspGlnLysGlyLeuThrAsnPheAsnLeuT 141
1387 CACTGCCATTTGATTC...TGCAATGTTGCGGTACAGCATCTGTAAT 1341
    |||
141 hIleProTyrGlySerGluCysAspValThrLeuThrAspLeu..... 155
1340 CCACGTGATTTTGTAAACAACAACACTGTTGTCATTTCTGTTTCA... 1294
    |||
156 ProLysHisArgTyrAlaGluThrThrValValLeuGlnAspAsnAlaAs 172
1293 ATTATTTGTTACCAAGTGCATGATGATGATGATGATGATGATGATGAT 1244
    |||
172 pLeuSerPheGlyLysThrThrArgLeuAsnHisValPheCysLeuTyr 189
1243 TGGAGCTGATTAACAGCTAGTAGACAGATGAGTATGATGATGATGAT 1194
    |||
189 hArgAsnValLysThr.....IleArgPheSerAspValSer 201
1193 ACTGCTTTTGA.....ACTCAATTCCTCCGATGCCAGATG 1156
    |||
202 AsnGlyHisGluValIleAlaSerThrGlyLysProLys..... 216
1155 CCGTTATGAATTTTG.....GATGGGACCAACCGGTGAC 1118
    |||
217 LysValGluMetLeuPheArgSerThrAspSerCysLysThrLeuGlnA 233
1117 CA.....GTTCATTTGCTATCATTTGTCACGCA... 1089
    |||

```

```

233 laalaargluasnlguphevalgiuphephelealeuenserproasp 249
1088 ...GTTTATCATTAATGACA.....TCGATTCGA 1060
250 SerAlaIyRHisGlyIleSerProLysGlyThrPheSerAspArgI 266
1059 AACCCCTAT.....ACTTCGCGGGGTGCC 1031
266 uasplleeralaproaspalaLysLysIleThrPhe.....ValGln 281
1030 ATTCTGCTTTGCGATGATGTAACGGTGAATCTGTAATTAATTAAT 981
281 lylglcysprovalasnlglymetasnasp..... 290
980 GCTGATGATGCTGCTTGAATTAATTTGCTAAATATTGGAATATTC 931
291 ..... 298
930 AACAGATTTAATGCTGCGCAAGACGTACGATATACAAATATGCGGATC 881
298 lasnaspclnilerpPheserLysPheArgThrPheArgPheGlyasn 315
880 GATACACCTTTCTATCATGCCAGATCACTATTACCATTAAGAACA 831
315 lnserrthvalphevahiscysglnvalglnvalcysleuLysLys... 330
830 AATGACCAATGTGTCACCAATGTCAGAACACACAGATTCGAGAC 781
331 ...AspLysSerLys...ThcysTyrLysLys..... 340
780 TGTTTAAACAGGTGTCGCGCAGCAAAACCTGTCAGCTGCGCACTTC 731
341 .....ValSerAspSerAsnLeuThrAlaGlnArgLeuA 352
730 GTTACTCAGAAAGATCTGCAGACCGAGAAATATCATGATGATGACA 681
352 rlphearghisllysArgSer.....IleThraspleuGlu 363
680 ACTGATATCAACACCTTGAATAGCATGATATCAACCTTGG 636
364 ArgArgThrThrArgSerAlaProThrAspAsnGlySerLeu 378

seq_name: SwtssProt_38: PGSL_BOVIN

seq_documentation_block:
ID PGSL_BOVIN STANDARD: PRT: 369 AA.
AC P21809: P79259:
DT 01-MAY-1991 (Rel. 18, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE BONE/CARTILAGE PROTEOGLYCAN I PRECURSOR (BIOGLYCAN) (LEUCINE-RICH PG I)
DE (PG-S1).
GN BGN.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-AORTA:
RX MEDLINE: 96113563.
RA XU J.H., RADIAKRISHNAMURTHY B., SRINIVASAN S.R., BERENSON G.S.;
RT "Primary structure of bovine aorta biglycan core protein deduced from
RT cloned cDNA.";
RL Biochem. Mol. Biol. Int. 37:263-272(1995).
RP
RP SEQUENCE OF 38-369.
RC TISSUE-CARTILAGE:
RX MEDLINE: 89255324.
RA NEAME P.J., CHOI H.U., ROSENBERG L.C.;
RT "The primary structure of the core protein of the small, leucine-rich
RT proteoglycan (PG I) from bovine articular cartilage.";
RL J. Biol. Chem. 264:8653-8661(1989).

```

```

RN [3]
RC SEQUENCE OF 38-63.
RP TISSUE-CARTILAGE:
RX MEDLINE: 89123388.
RA CHOI H.U., JOHNSON T.L., PAL S., TANG L.H., ROSENBERG L.C.,
RA NEAME P.J.;
RT "Characterization of the dermatan sulfate proteoglycans, DS-PGI and
RT DS-PGII, from bovine articular cartilage and skin isolated by occyl-
RT sepharose chromatography.";
RL J. Biol. Chem. 264:2876-2884(1989).
CC -1- TISSUE SPECIFICITY: FOUND IN THE EXTRACELLULAR MATRICES OF SEVERAL
CC CONNECTIVE TISSUES, SPECIALLY IN ARTICULAR CARTILAGES.
CC -1- PTM: THE TWO GLYCOSAMINOGLYCAN CHAINS ATTACHED TO BIGLYCAN CAN BE
CC EITHER CHONDROITIN SULFATE OR DERMATAN SULFATE.
CC -1- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC FAMILY.
CC -1- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC MANY PROTEINS. NUMBER IN THIS PROTEIN: 10.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: S82652; AAB46746.1; -
DR PIR: A33701; A33701.
DR PFAM: PR00560; LRR: 8.
DR PFAM: PR01462; LRRNT: 1.
KW Glycoprotein; Connective tissue; Extracellular matrix; Proteoglycan;
KW Signal; Repeat; Leucine-repeat.
FT SIGNAL 1 19
FT PROPEP 20 37
FT CHAIN 38 369
FT DOMAIN 93 316 BONE/CARTILAGE PROTEOGLYCAN I.
FT REPEAT 93 106 LEUCINE-RICH REPEATS.
FT REPEAT 117 130 LRR 1.
FT REPEAT 141 154 LRR 2.
FT REPEAT 162 175 LRR 3.
FT REPEAT 186 199 LRR 4.
FT REPEAT 211 224 LRR 5.
FT REPEAT 232 245 LRR 6.
FT REPEAT 256 269 LRR 7.
FT REPEAT 280 288 LRR 8.
FT REPEAT 303 316 LRR 9.
FT REPEAT 342 42 LRR 10.
FT CARBOHYD 48 48 GLYCOSAMINOGLYCAN.
FT CARBOHYD 271 271 GLYCOSAMINOGLYCAN.
FT CARBOHYD 312 312
FT DISULFID 64 77
FT DISULFID 322 355
FT CONFLICT 152 152
FT CONFLICT 188 188
FT CONFLICT 354 354
FT CONFLICT 368 369
SO SEQUENCE 369 AA; 41509 MW; F1CC673B CRC32;

alignment_scores:
Quality: 106.50 Length: 299
Ratio: 0.873 Gaps: 13
Percent Similarity: 40.803 Percent Identity: 20.736

alignment_block:
US-09-323-427-2/rev x PGSL_BOVIN ..
Align seg 1/1 to: PGSL_BOVIN from: 1 to: 369
775 AAACAGGTGTCGCGCAGCAAAACCTGTCGACGTCGCGCACTTGTGTTA 726
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
87 LysGlnIleSerProAspThrThrLeuAsnAspLeuGlnAsnAsnAspI 103

```

```

725 CTCAC.....AGAAAAGAT..... 713
103 eSergIuleuArgLysAspAspPheLysGlyLeuGlnHisLeuTyrAlaL 120
712 .....CTGCAGAACCGGAGATATCATGTAGTACGAACTGATATCAACA 668
120 euValLeuValAsnAsnLysIleSerLysIleHisGlyLysAlaPheSer 136
667 CCCTGAAATTAGCATGATGAATCAAGCTTTGCCAGTTGATTACGTCAC 618
137 ProLeuArg.....LysLeuGlnLysLeuTyrIleSerLys 148
617 CGTGCAGCTTTCGACATATATGACAGAACCTGATATCTCGTGCAGTACA 568
148 sAsnHisLeuCysGlyIleProProAsn..... 157
567 AATGGAATGTCATGTCACCATTTGCTTCTCAATGTTATGGGTTAA 518
157 ..... 157
517 GCATTGCATGATGTCGCCCTCATTTATACCATTTGCTTAAATTTGCT 468
158 .....LeuProSerSerLeu.....Va 163
467 CCAATTCAGAACGATTAATAATATGTTAGA.....ATCA 433
163 IGIleuArgIleHisAspAsnArgIleArgLysValProLysGlyValP 180
432 TCGAAGCATA..... 422
180 heserGlyLeuArgAsnMetAsnCysIleGluMetGlyGlyAsnProLeu 196
421 .....ATMAAATGCCATA 408
197 GluAsnserGlyPheGluProGlyAlaPheAspGlyLeuLysLeuAsnTy 213
407 TATATTCGTTTCTTCTTATCATCCTTCTAATACATAATTTTGTGCTAACAA 358
213 rLeuArgIleSerGluAlaLys.....LeuThrG 223
357 AATATATAGTATGAGGAATTAATTAAGTAAATCAATAGTATGATTTTC 308
223 LysLe.....ProLysAspLeuProGluThrLeuAsnGlyLeuHis 236
307 ATCAAACTTCTTCTATCGCTTTATAGCTTTCGAAGAAGCTTATTCATTA 258
237 LeuAspHisAsnLysIleGlnAlaIleGluLeuGluAspLeuLeuArgTy 253
257 TTCAGTAATCTTTATATGATCATCTATTTAAATGTTTCATCATTTAGGCC 208
253 rSerLysLeuTyrArgLeu.....GlyLeuGln 263
207 ATGAAATGTTGCTTGTATATCA...TCATTTCAACTGTGCT... 164
263 IsAsnGlnIleArgMetIleGluAsnGlySerLeuSerPheLeuProThr 279
164 ..... 164
280 LeuArgGlyLeuHisLeuAspAsnAsnLysLeuSerArgValProAlaGly 296
163 .....ATTATATCTAAGCATTTATATCATCTTTGATGATTAATA 130
296 yLeuProAspLeuLysLeuLeuGlnValValTyrLeuHisThrAsnAsn 313
129 TCACAAATTTATACCTTG.....TATTCGCCAATTTTATGGGCTC 89
313 IeThrLysValGlyValAsnAspPheCysProValGlyPheGlyVal 328
seq_name: swissProt_38:KM65_YEAST
seq_documentation_block:
ID KM65_YEAST STANDARD; PRT; 742 AA.

```

```

AC 003656;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PROBABLE SERINE/THREONINE-PROTEIN KINASE YMR216C (EC 2.7.1.-).
GN YMR216C OR YMR261.10C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA DEEDMAN K., BROWN D., BOWMAN S., BARRELL B.G., RAJANDREAM M.A.,
RA WALSH S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
CC PROTEIN KINASES. SIMILAR TO S.POMBE DSK1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: Z49809; CAA89931.1; -
CC DR HSSP: P24941; 1A01.
CC DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
CC DR PROSITE: PS00109; PROTEIN_KINASE_DOM; 1.
CC DR PFAM: PF00069; pkinase; 3.
CC KW Hypothetical protein; transferase; Serine/threonine-protein kinase;
CC ATP-binding.
CC FT DOMAIN 158 706 PROTEIN KINASE.
CC FT NP_BIND 164 172 ATP (BY SIMILARITY).
CC FT BINDING 187 187 ATP (BY SIMILARITY).
CC FT ACT_SITE 294 294 BY SIMILARITY.
CC SO SEQUENCE 742 AA; 83237 MW; DCF32417 CRC32;

alignment_scores:
    Quality: 104.50      Length: 303
    Ratio: 0.816        Gaps: 13
    Percent Similarity: 42.244      Percent Identity: 20.462

alignment_block:
US-09-323-427-2/rev x KM65_YEAST ..
Align seg 1/1 to: KM65_YEAST from: 1 to: 742

1737 AGCAACACACAAACAAACAAACAAACAAACAAACAAACAAACAAATATA 1688
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
419 SerGlyAsnArgAsnIleProSerSerIleAsnAsnSerIleAsnAs 435
1687 CCCCATCAAGTGGAGAGAAGACAGACAGACATCTTAAGTTTCTVAAA 1638
1  |||  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
435 nglyIleGlyIleLysAsnSerAsnAsnSerPheLeuAsnSerValProH 452
1637 ATCGAATTTACTAAATCTTCGAATAGATGATGCTTCTTAATGCTTCTG 1588
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
452 isSerValThrArgMetPheIleAsnGluAspSerAsn..... 464
1587 TACTACACTTATTCGATTTGCTTATTCGATTCGGTTGACAAATGCTGCG 1538
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
465 .....AspAsnAsn..... 467
1537 AAGTGACCCAGAAATGATGACCAACTTCAATATACAAATCAATTTT 1488
|||||  :|||  |||  |||  |||  |||  |||  |||  |||  |||
468 .....AsnAsnAspAsnSerLysAsnLysAsnAsnAsn..... 480
1487 AATACAGTATGATTTGGAAGGACATGTTATGTGAAGGCTTTATGA 1438
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:

```



```

481 AsnSerAsnAsnAsnAsnGluAspIleMetAsnThrProLeuHisG1 497
1437 TCACAGAGGTTGCCGTAATGATGAGGTGACGTCAGTGGCCGGAATT 1388
      :::::|||||
497 uGlu.....GlnLeuAlaAspSerL 504
1387 CACTTCCATTGATGATGCAATGTGGCGGTACGATCTGTGATCA 1338
      ||||| ||||| ::::: |||
504 euserThpAspIleSerAsnIleSerGlnSerSerAspThrAsn... 519
1337 CGTGTATTGTTGTACACAACTGTCTGATTTGTTGTTATCCATTAT 1288
      ||| ::::: |||
520 ...GlyProTyrIleSerAsnThr..... 526
1287 TGTATACCAAGTGTGATGATATGATGATCAATGTTTACATGAAG 1238
      |||||
527 .....MetAspS 529
1237 CTGATTAACAGTATGTCGA.....CAG 1215
      ::::: |||||
529 eAsnSerAsnValSerThrAspIleAsnSerProGluAsnLeuIleGln 545
1214 ATTGAGTATCTGAATC.....AC 1195
      ||||| ::::: |||
546 IleYsIleAlaAspLeuGlyAsnAlaCysTrpTyrAspGluHisTyrTh 562
1194 AACTGCTTTTCAACCTCAAAATGTCGCCGATCCAGTATCCGTTATGA 1145
      ::::: ||||| |||
562 rAsnSerIleGlnThrArgGluTyrArgSerPro.....GluV 575
1144 TTTTGGAGTGGAGCCACCGCTCACACAGTCAATTGCTATCATTTG 1095
      ::::: |||||
575 alLeuLeuGlyAlaPro..... 580
1094 CAGCAGATTTCATTAATGACATGCGATTCGAAACCGTTGATACTTT 1045
      ||| ||| ::::: |||
581 .....TrpGlyCysGlyAlaAspIleTrpSerThrAl 591
1044 CTCGCCGCTGTC..... 1032
      |||
591 acYsLeuIleheGluLeuIleThrGlyAspHeuLeuPheGluProAspG 608
1031 ....CATCTGCTTGTGATGATGATGATAC...GGTGATCTGTGAA 990
      ||||| ||||| ::::: |||
608 lueGlyHisSerTyrThrLysAspAspHisIleAlaGlnIleLeu 624
989 ATTCTAAATGCTGATGATGATGCTTGTGATAATATTGCTAAATAATT 940
      ::::: |||||
625 LeuLeuGly.....GluLeuProSerTyrLeuLeuAlaGlnG1 637
939 GGAATAT 933
      ::::: |||
637 yLysTyr 639
seq_name: SwissProt_38:APMU_PIG

```

```

RT glycosylated domain."
RL J. Biol. Chem. 266:9678-9686(1991).
RN [2]
RP SEQUENCE OF 1-503 FROM N.A.
RC TISSUE-SUBMAXILLARY GLAND.
RX MEDLINE: 88087170.
RA TIMPTE C.S., ECKHARDT A.E., ABERNETHY J.L., HILL R.L.;
RT "Porcine submaxillary gland apomucin contains tandemly repeated,
RL identical sequences of 81 residues.";
RN J. Biol. Chem. 263:1081-1086(1988).
RN [3]
RP SEQUENCE OF 45-80.
RC TISSUE-SUBMAXILLARY GLAND.
RX MEDLINE: 87280230.
RA ECKHARDT A.E., TIMPTE C.S., ABERNETHY J.L., TOWMADE A.,
RT JOHNSON W.C. JR., HILL R.L.;
RL "Structural properties of porcine submaxillary gland apomucin.";
RN J. Biol. Chem. 262:11339-11344(1987).
RN [4]
RP CARBOHYDRATE-BINDING SITES, AND SEQUENCE OF 45-125.
RC TISSUE-SUBMAXILLARY GLAND.
RX MEDLINE: 97248516.
RA GERKEN T.A., OMENS C.L., PASUMARTY M.;
RT "Determination of the site-specific O-glycosylation pattern of the
RT porcine submaxillary mucin tandem repeat glycopeptide. Model proposed
RL for the polypeptide:galnac transferase peptide binding site.";
RN J. Biol. Chem. 272:9709-9719(1997).
CC -1- FUNCTION: APOMUCIN IS PART OF MUCIN, THE MAJOR GLYCOPROTEIN
CC SYNTHESIZED AND SECRETED BY MUCOUS CELLS OF THE SUBMAXILLARY
CC GLAND. ITS HIGHLY VISCOUS AQUEOUS SOLUTIONS SERVE TO LUBRICATE
CC THE ORAL CAVITY AND TO PROTECT IT FROM THE EXTERNAL
CC ENVIRONMENT.
CC -1- SUBUNIT: INTERMOLECULAR DISULFIDE BONDS COULD HELP MAINTAIN A
CC MULTIMERIC MUCIN STRUCTURE.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: SUBMAXILLARY MUCOSAE.
CC -1- DOMAIN: CONTAINS TANDEMLY REPEATED, IDENTICAL SEQUENCES OF 81
CC RESIDUES.
CC -1- PTM: EXTENSIVELY O-LINKED GLYCOSYLATED ON SER AND THR RESIDUES OF
CC THE REPEAT UNITS. HIGHEST GLYCOSYLATION APPEARS TO OCCUR ON SER
CC RESIDUES WHICH HAVE GLY AT POSITIONS AT +2 OR -2 FROM THE
CC GLYCOSTYLATION SITE OR, WHERE GLY IS THE PENULTIMATE RESIDUE, THE
CC PRESENCE OF PROLINE (USUALLY AT POSITION +3 OR -3) APPEARS TO ALSO
CC ENHANCE GLYCOSYLATION.
CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M61883; AAA30998.1; -
DR EMBL: M21174; AAA30990.1; -
DR PIR: A40009; A40009
DR PROSITE: PS01185; CTCK_1; 1.
DR PROSITE: PS01225; CTCK_2; 1.
DR PROSITE: PS01208; WMFC; 1.
DR PFAM: PF00007; Cys_knot; 1.
KW Glycoprotein; Repeat.
FT DOMAIN 1
FT REPEAT 1
FT REPEAT <1 368 81 AA TANDEM REPEATS.
FT REPEAT <1 44 1 (FRAGMENT).
FT REPEAT 45 125 2.
FT REPEAT 126 206 3.
FT REPEAT 207 287 4.
FT REPEAT 288 368 5.
FT REPEAT 369 391 6 (INCOMPLETE).
FT DOMAIN 929 995 WMFC.
FT DOMAIN 1062 1146 CTCK.

```



```

419 uHISValGlyaspProIleValValGlyAsnthrPheGlyArgVal.... 434
961 ATAAATATTTGCTAAATTAATTG..... 939
435 ..ArgAlaMetValAsnAspIleGlyArgValGlyThrIleGlyPro 450
938 GAATATCCACACAGATTTA.....ATGGCTGGCCA 910
451 SerthrProValGluIleThrnglyLeuAsnAspValProGlnIleGly 467
909 AGAAGCTCAGCTATACAAATATGGGATCGATCAGACGTTTTCATCAT 860
467 pGlnPheLeuValPheLysAspClyuThrAlaArgSerValGlyGluA 484
859 GCCAGATCAGTATTATCCATTAAAGAACCAATAGCAATGTGTGCCACCA 810
484 laArgAlaSerLysGlnLeuGlnGlnIleArgSerAsp....LysAla 498
809 CAATGTTTCAGAACCCACAGAGATTGGAGCTGTTAAACAGTGT..... 765
499 LysLeuSerLeuAspIleuPheGlnGlnIleLysGlnGlyAspVally 515
764 .....GCCGACGAAACCTGCTGCAGCTGCC 737
515 sAspIleAsnLeuIleValLysAlaAspValGlnGlySerAlaGluAla 532
736 AACTGTGTTTACTCAGAAAGATCTGCAGAACCGAGAAATATCATGAT 687
532 eurThrAlaAlaLeuGlnLysIleGlyValGlnGlyVallyLysIle 548
686 GTACGACACTGATATCAACACCCCTTGAATATGAGATGAATCAAGCTTT 637
549 IleHisThrGlyValGlyAlaIleThrGlnSerAsp..... 560
636 GCCAGTGAATTTACGTCACCGTCACCTTGCACACATTAATGACAACTG 587
561 .....I 561
586 TAATCTTGCTGTCAGTACAAATGGAATGCAATGCAACCATTTGGCTTC 537
561 leIleuAlaSerAlaSerAsnAlaIleValIle.....GlyPhe 574
536 TCATATG 531
575 AsnVal 576

seq_name: SwissProt_38:MUCS_BOVIN
seq_documentation_block:
ID MUCS_BOVIN STANDARD; PRT; 563 AA.
AC P98091;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SUBMAXILLARY MUCIN-LIKE PROTEIN.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SUBMAXILLARY GLAND;
RX MEDLINE: 90370871.
RA BHARGAVA A.K., MOITACH J.T., DAVIDSON E.A., BHAVANANDAN V.P.;
RT "Cloning and cDNA sequence of a bovine submaxillary gland mucin-like
RT protein containing two distinct domains";
RL Proc. Natl. Acad. Sci. U.S.A. 87:6798-6802(1990).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: SUBMAXILLARY MUCOSAE.
CC -1- SIMILARITY: TO PORCINE APMUCIN.
CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTEINE KNOT-LIKE DOMAIN (CTCK)
CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTEINE KNOT-LIKE DOMAIN (CTCK)

```

1428 TTGCCGTAATGATGAAGGTGGACGTCAAGTTGCCGGAATTTCACCTTCCAT 137

01-JAN-1990 (Rel. 13, Created)


```

RP SEQUENCE FROM N.A.
RX MEDLINE: 92333688.
RA ALBRECHT J.-C., NICHOLAS J., BILLER D., CAMERON K.R., BIESINGER B.,
RA NEWMAN C., WITTMANN S., CRAXTON M.A., COLEMAN H., FLECKENSTEIN B.,
RA HONESS R.W.;
RT "Primary structure of the herpesvirus saimiri genome.";
RL J. Virol. 66:5047-5058(1992).
RN
RN SEQUENCE FROM N.A.
RX MEDLINE: 89036162.
RA GOMPELS U.A., CRAXTON M.A., HONESS R.W.;
RT "Conservation of the gH gene from herpesvirus saimiri.";
RL J. Gen. Virol. 69:2819-2829(1988).
RN
RP SEQUENCE OF 1-60 FROM N.A.
RX MEDLINE: 90063548.
RA HONESS R.W., CRAXTON M.A., WILLIAMS L., GOMPELS U.A.;
RT "A comparative analysis of the sequence of the thymidine kinase gene
of a gammaherpesvirus, herpesvirus saimiri.";
RL J. Gen. Virol. 70:3003-3013(1989).
CC -1- FUNCTION: GH AND GL FORM A COMPLEX THAT INDUCES NEUTRALIZING AND
CC SYNCYTIA INHIBITING ANTIBODIES (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEIN H.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X64346; CAA45645.1; -
DR EMBL: D00400; BAA00302.1; -
DR EMBL: D00543; BAA00433.1; -
DR PIR: JQ0010; VGBELL.
KW Glycoprotein, Transmembrane, Signal.
FT SIGNAL 1 16
FT CHAIN 17 717 GLYCOPROTEIN H.
FT CARBOHYD 43 43 POTENTIAL.
FT CARBOHYD 59 59 POTENTIAL.
FT CARBOHYD 80 80 POTENTIAL.
FT CARBOHYD 128 128 POTENTIAL.
FT CARBOHYD 444 444 POTENTIAL.
FT CARBOHYD 560 560 POTENTIAL.
FT CARBOHYD 613 613 POTENTIAL.
FT CARBOHYD 675 675 POTENTIAL.
FT SEQUENCE 717 AA; 82583 MW; 517E5EA CRC32;
SO

```

```

alignment_scores:
  quality: 99.50      length: 660
  ratio: 0.358        gaps: 39
  Percent Similarity: 42.121  Percent Identity: 20.000

```

alignment_block:

US-09-323-427-2/rev x VGLH_HSVSA ..

Align seg 1/1 to: VGLH_HSVSA from: 1 to: 717

```

1706 ACAACACAGCATATATACCCCATCAAGTGAGAGAGACAGACAGCAGC 1657
182 ThrileSerCysHisIleThrProAsn..... 190
1656 AATCTAGTTTCTTAAATAAGAAATTAATCTTCTGAATGAGA 1607
191 ....TyrLeuPheValSerValGluPheThrLysPheAspSerLeu... 204
1606 TTCGCTTATTCCTTCTGCTACTACACTATATGCAATGCTTATTCGATT 1557
205 .....ThrLeuLeuPheIleHisSerHisTyrLeu 214

```

```

1556 CCG.....GTTGACAAATGCTGTGCGAAGTGCAGCAGCA 1525
1215 ProProLeuLysGlyHisIleValIleTyrAsnAspIleGluGlyAlaSerAs 231
1524 AATGATGATGACCAACTTCATATACATCAATTTTATACACGATATG 1475
231 n.....AspValPheSerLeuValIle...PheSerThrTyrAspL 244
1474 CATTCGAAGACATGTT..... 1458
244 euPheGlyLysHisValGluSerPheLysPheAspIleAlaLysValPhe 260
1457 .....TATGTGAAGGCTTTTANGA 1438
261 ArgGluIleIleGluThrProProLeuThrPheIleLysAsnLeuGlnAs 277
1437 TCAGAAGCTTGCCCTAATGATGACAGTGCAGCTCAAGTTGCCCGCATTT 1388
277 P.....GluMetPheThrIleG 283
1387 CACTTCATTTGATTCATGCAATGTTGCCGCTACAGCATCTGTAATCCA 1338
283 IuleArg...AspGlyCysAsnIle...AsnAsnIleValAsnPro 296
1337 CGTGGT...ATTTTGAACAACAACAGTGTGTCATTTGCTTCATCCATT 1291
297 LysThrPheLeuPheAlaPheLysAlaValAlaIleHisPheLeuValI 313
1290 ATTGTGTACCAAGTGTATGCTGCATATCGAGTCAATGCTTT.....T 1247
313 eaSpSerLeuArgThrGlnGlnHisIleLeuLeuAsnCysPheLysAsnT 330
1246 ACATGGAAGCTGTATAAACAGTTACTGACACAGATTGACGTTTCGAATC 1197
330 yrmel.....SerGluLeu 334
1196 ACAATGCTTTTCACAACTCAATTTGCCGATGC..... 1163
335 GluPheLeuArgLysLeu...MetGluSerCysPheGluPhePheGluP 350
1162 .....CAGTATGCCGCTTATGAATTTTGGATGCTGGACCAACCGGTC 1121
350 heaSpPheProThrValIleGluThrLeuAlaIleSerGlnAlaLeu 366
1120 AAC.....CAGTTCAA 1110
367 AsnValProLysHisValIleThrSerLeuSerHisGluAspSerThrAs 383
1109 TTTGCTATGCTGTCAGCCAGTTTATCATTAATGACATGCGATTCTGA 1060
383 nmetLeuSerLeu.....PheArgLeuSerArgHis..... 393
1059 AACGCTGATACCTTCTGCCCGCTGTCATTCCTGCTTGTGCGATGATG 1010
394 .....SerLysHisValSerAlaValAlaIleSerGluIle 405
1009 GTAACGCTGATCTGCGAATTAATGCTGATGA..... 972
406 IleAspLeuIle...SerHisIleTyrThrAlaTyrSerTyrThrTyrMet 421
971 .....TGTGCTTGTGATAAATTTTG...CTAA 946
422 LeuThrSerSerGluArgLysMetLeuLeuAspAlaTyrIleValLeuAs 438
945 TAAATTTGAATATCCACAGATTTAATGCTGCGCCACAGATCAGCTAT 896
438 nasPleuMetHisLysAsnGluThrValLysLysGlnAspLeuLeuProt 455
895 ACAAAATATGCG..... 885
455 yrfValLeuSerSerSerMetCysThrSerLeuGluIleGlyAsnLeuLeu 471
884 .....GATGATCACACGCTTTTCTATCA 862

```



```

472 LeuHisphenylglycylaspvalleuaspvaltyrGluThrPheSerPr 488
861 ATGCCAGATCGATTACCAT.....AAGACCAATATAGCGAT 821
488 ocYstyLeuSerLeuArgPheAspPheThrLysLeuIleThr 504
820 GTGTCGACAC..... 809
505 GluPheProGlnSerSerLeuIleAlaGlnLysGluIleAsnLeuGlyTh 521
808 .....AATGTCAGAACACAGCATTCGAGC 781
521 rasnGlyPhePheGlnThrLeuHisMetArgHisIleThrSerLeuGluI 538
780 TGTAAACAGGTG..... 767
538 IeLeuProIleIleLysCysIleLysSerLeuSerThrAspIleIleLeu 554
766 ..GTGCCGAGCAAAACCTGCTGACGCTGCGCACTTCGTTACTCAAG 720
555 SerIleProLeuLysAsnIle.....ThrTyraIleIleSerTh 567
719 AAAAGATCTGCAGACCGGAGAAATATCATTTGATGACGATATCA 670
567 rlyspProValProAsnSerLysIle.....TyraSpValSerGluV 581
669 CACCCCTGAATTAGCATGATATCAAGCTTCCAGTTGATTACTGTC 620
581 alheLeuLysThrSerMetIleIleSerAlaValAsn..... 593
619 ACCGTCACCTTCGCAACATTAATGACACCTGTAATCTTGTCAGTA 570
594 .....AsnAspCysLys.....ProTy 599
569 CAATAAGGATCTGCATGATGCACATTTGGCTTCGAATGTTATGGGTT 520
599 rGlnGlyGlySerAlaAlaHisGln..... 607
519 AAGCATTCATTCATTGCTGCGTCATTTACCATTTGCTTAATTC 470
608 ..... 614
469 GTCCCAATCAGAGCGCATAAATAATGTAGATCATCGAACCAATAT 420
615 ThrValProArgArg..... 619
419 AAAACCTGCATATATA.....TTCGTTCTCTTAATCATCCTCTTAATA 376
620 .GlyCysProTyrcysSerSerValValIleuSerTyraSpGluSerGln 636
375 ACTAATTTAGCTAACAAATATATAGTAGAGAAATAATATCTGTAAT 326
636 Ly.....PheGlnSerMetMetIleThrAspThrTyraGln 649
325 ACAATAAGATATTTTTCATCAAAATCTTCATCGCTTTTATACCTTC 276
650 .....GluAsnLeuPhe.....ThrGluHisSerProPheheLeuLys 662
275 TGAAGAAGCTT...ATTCAATATTCAGTAATCTTTATATATGACATAGTG 229
662 polysnLeuHisIleHisTyLeuIleLeuMetAsnAsnGlyThrValI 679
228 TAAATGTTTCATCAT...AGGCAATGAATAGTTTCGTTTGTATATATC 183
679 IeGluValArgGlyAlaTyraGlnAlaArgLeuValAsnPheIleIleVal 695
182 ATCATATTAACATCTGCTATTTTATTTACAGCATTTATCATTTGGANA 133
696 IleMetVal.....PheIleLeuPheLeuValGlyLeuTyrlle 708
132 ATATCACAATATATACCTTGATGTG 108

```

```

708 uLeuTyrlsLeuPheValTyrlleu 716
seq_name: SwissProt_38:YAY3_SCHPO
seq_documentation_block:
ID YAY3_SCHPO STANDARD; PRT; 649 AA.
AC Q10211;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 74.5 KD PROTEIN C4H3.03C IN CHROMOSOME I.
GN SPAC4H3.03C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA MORPHY L., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.
CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: 269380; CAA93342.1;
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 265 285 POTENTIAL.
FT TRANSMEM 564 584 POTENTIAL.
FT TRANSMEM 626 646 POTENTIAL.
SQ SEQUENCE 649 AA: 74488 MW: F04F8763 CRC32;

alignment_scores:
Quality: 97.50 Length: 403
Ratio: 0.554 Gaps: 23
Percent Similarity: 43.672 Percent Identity: 22.333

alignment_block:
us-09-323-427-2/rev x YAY3_SCHPO ..
Align seg 1/1 to: YAY3_SCHPO from: 1 to: 649
1541 GTCCGAAGGTGAGCCAGAAATTTGAATGTGACCAACTTCATTAACAAATCA 1492
36 LeuAspGlySerValGluMetMetCysTrpPro.....As 47
1491 TTTTATACACGTAATGCATTC.....GAAGGAC 1463
47 nPheAspSerProSerIlePheHisAlaArgIleLeuAspAlaArgAlaGln 64
1462 AGTTTATGTGAAGGCTTTATGATCAGAAAGGTGCGGTATGATGA 1413
64 isPheSerIleThrProIle...GluGlnThrSerCysLys..... 76
1412 GGTGACGCTCAAGTTCGCCGAATTCCTTCATTTGATTCATTCGAATGT 1363
77 .....GlnMetTyrlGluProSerThrAsnI 85
1362 TGGCGGTACAGATCTCGAATCCAGCGTGATTTTGTGAACAACAACCTG 1313
85 eLeuHisThrLysPheTyrlSerGluArgGlyValLeu.....ArgL 99
1312 TTGTCATTTGCTTCAT.....CCATTATTTGTT 1284
99 euleuAspPhePheHisArgProTirpGluAspTyrlGluProLeuTyrlPro 115

```

```

1283 ACCAAGTTGATCGTCA.....TATCAGT 1258
      :: |||::
116 TrpLeuLeuArgValSerCysIleArgGlyThrSerArgIleLeuSLe 132
      :: |||::
1257 ACAATGCTTTACATGGAAGCGATGAACAGATTAGTCACAGATTGAG 1208
      :: |||::
132 uGluCysPheProAlaLeuAspTyrAlaArgIleSerHisGluThrArgV 149
      :: |||::
1207 TATTCGAATCACA...ACTGCTTTCAACCTCAAAATTCCTCCGATGCA 1161
      :: |||::
149 aSerLysIleThrGluAsnTyrTyrGlnAlaGluPheValProAlaSer 165
      :: |||::
1160 GATGCGCTTAAGAATTTTGATGTGTGACCAACCGCTCA..... 1119
      :: |||::
166 GlyAspProLysTyrIleLeuAspCysValProSerGlyAspIleLeu 182
      :: |||::
1118 .....CCAGTTCATTTGCTATCATTTGCTGACG 1091
      :: |||::
182 sIleAspLeuGluLeuIleTyrProAlaGlnHisLeuIleGluGlyG 199
      :: |||::
1090 CAGTTTATCAATAATGACATGCGATTCGAAACCGTTGATCTTCTGC 1041
      :: |||::
199 IyAlIleSerTyrLeuGluLeuGluGluGlnGluIleThrPhe... 214
      :: |||::
1040 GCGGTTCCATTCCTGCTTGTGATGATGTAACGATGATCTGTGGA 991
      :: |||::
215 .....IlePheArgGlnGluGlyLeuGlyProAsnValAs 226
      :: |||::
990 AATCTTAATGCTGATGATGCTGCTTGTGATAAATTTGCTAATAAT 941
      :: |||::
226 PTrValIleThrProAsn.....LeuValAspLysL 236
      :: |||::
940 TGAATATCAACAGATTAAATGCTGCGCAAGACGACGATATACAAA 891
      :: |||::
236 euGluAspSerThrLysArgTyrTrpArgAlaTrpIleGlnGlnCysVal 252
      :: |||::
890 TATGCGGATGATCACAGCTTTCTATCA.....TCCAGATCAG 850
      :: |||::
253 TyrThrGlyArgTyrArgGluPheValGlnArgAsnAlaLeuThrLeu 269
      :: |||::
849 TATTAACATTAAGAACCAAAATAGCGAATGTGT...CGACCACATGTT 803
      :: |||::
269 sIeuLeuIleTyrGluProThrGlyValAlaIleAlaSerProThrPheS 286
      :: |||::
802 CAGAACCAACAGATTGCGAGCTGTAACAA..... 771
      :: |||::
286 eTrLeuProGluAspLeuGlyGlyValAlaArgsnTrpAspTyrArgPheThr 302
      :: |||::
770 .....GGTGGTCCGCGACCAAAACCTGCTGACGCTGCCAATT... 732
      :: |||::
303 TrpIleArgAspSerAlaPheThrIleTyrAlaLeuAlaGlnLeuGlyPh 319
      :: |||::
731 .....CGTTTACTCAGA 719
      :: |||::
319 eArgAlaGluAlaValGluTyrMetSerPheIleTyrHisAlaLeuLysL 336
      :: |||::
718 AAAGATCTGCAGAACCGAG...AATATCATTTGATGTACGAATGATATC 672
      :: |||::
336 yLysAsnLysAspGlyGlyIleAsnIleVal..... 346
      :: |||::
671 AACACCCCTTGAATTAAGCGATTAATCAAGCTTTG...CCAGTTGATT 625
      :: |||::
347 .....TyrSerIleHisGlyAspSerGlnAsnLeuGluGlnValGlnLe 361
      :: |||::
624 ACGTCAAC...CGTGACCTTGCACAACATTAATGAGCAACCTGTAACTGTG 578
      :: |||::
361 uThrIleLeuArgGly.....TyrTyrAsnSerHisProValArgIle. 375
      :: |||::
577 CTGAGTACAAAATGGAATCTGCATCTGCACATTTGGCTTCTCAATGTT 528
      :: |||::
376 .....Gly AsnAlaIleValHisHisLeuGln. LeuAspIleT 388
      :: |||::
527 ATGGG 523

```

```

|||||
388 yRGly 389

```

```

seq_name: SwissProt_38.ycf2_MARPO

```

```

seq_documentation_block:

```

```

ID ycf2_MARPO STANDARD; PRT; 2136 AA.

```

```

AC P09975;

```

```

DT 01-MAR-1989 (Rel. 10, Created)

```

```

DT 01-MAR-1989 (Rel. 10, Last sequence update)

```

```

DT 01-JUL-1993 (Rel. 26, Last annotation update)

```

```

DE HYPOTHEtical 259 KD PROTEIN (ORF 2136).

```

```

GN ycf2.

```

```

OS Marchantia polymorpha (Liverwort).

```

```

OG Chloroplast.

```

```

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiopsida;

```

```

OC Marchantiales; Marchantiaceae; Marchantia.

```

```

OC [1]

```

```

RP SEQUENCE FROM N.A.

```

```

RN RA OHYAMA K.;

```

```

RL Submitted (OCF-1986) to the EMBL/Genbank/DBJ databases.

```

```

RN [2]

```

```

RP COMPLETE GENOME.

```

```

RA OHYAMA K., FUKUZAWA H., KOHCHI T., SHIRAI H., SANO T., SANO S.,

```

```

RA UMESONO K., SHIKI Y., TAKEUCHI M., CHANG Z., AOTA S., INOKUCHI H.,

```

```

RA OZERI H.;

```

```

RT "Chloroplast gene organization deduced from complete sequence of

```

```

RT Liverwort Marchantia polymorpha chloroplast DNA."

```

```

RL Nature 322:572-574(1986).

```

```

CC -1- SIMILARITY: NO A SIMILAR ORF IN OTHER PLANTS CHLOROPLAST.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

```

```

CC at the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).

```

```

CC

```

```

CC EMBL: X04465; CAA28078.1;

```

```

DR PIR: A05037; A05037.

```

```

DR PIR: S01591; S01591.

```

```

DR MENDEL: S295; MARPO.ycf2.1.

```

```

DR PFAM: PF00004; AAB.1.

```

```

DR Chloroplast; Hypothetical protein.

```

```

KW SEQUENCE 2136 AA; 259911 MW; A2FB8B07 CRC32;

```

```

SQ

```

```

alignment_scores:

```

```

Quality: 97.00 Length: 662

```

```

Ratio: 0.316 Gaps: 33

```

```

Percent Similarity: 46.375 Percent Identity: 18.278

```

```

alignment_block:

```

```

US-09-323-427-2/rev x ycf2_MARPO

```

```

Align seg 1/1 to: ycf2_MARPO from: 1 to: 2136

```

```

1731 AACACAAA.....CAACAACACAAACAC.....AA 1703

```

```

709 AsnAsnLysLeuIleThrTrpLysLysIleSerAsnLysLeuValIleSe 725

```

```

1702 CAACAGCAATAATACCCATCAACAGTGGAGAGAGAGACAGACAAATC 1653

```

```

725 rAsnSerGluTyrAsnLysLysIleIleTrpAsnLysLysAsnMetLysPheP 742

```

```

1652 TTAGTTTCTTAATAATGAAATTACTAATCTTCTGAATGATGATTG 1603

```

```

742 heSerPheSerLysAsnSerValIleAspThrPhePheAsnLys 758

```

```

1602 TCTTATGCTTCTGCTACTACATTATGCAATTGCTTATTTGATTCGG 1553

```

```

759 SerPheAsnIleIle.....ThrValIlePheAspLys L 770

```



```

seq_documentation_block:
ID   GALU_PSEAE          STANDARD;          PRT:          279 AA.
AC   059633:
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   UTP-GLUCOSE-1-PHOSPHATE URIDYLTRANSFERASE (EC 2.7.7.9) (UDP-GLUCOSE
DE   PYROPHOSPHORYLASE) (UDPGP) (ALPHA-D-GLUCOSYL-1-PHOSPHATE
DE   URIDYLTRANSFERASE) (URIDINE DIPHOSPHOGLUCOSE PYROPHOSPHORYLASE).
GN   GALU.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC   Pseudomonas.
CC   [1]
CC   SEQUENCE FROM N.A.
CC   STRAIN-ATCC 10145;
CC   CHANG H., LEE C., PENG H.;
CC   Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
CC   -1- FUNCTION: MAY PLAY A ROLE IN STATIONARY PHASE SURVIVAL (BY
CC   SIMILARITY).
CC   -1- CATALYTIC ACTIVITY: UTP + ALPHA-D-GLUCOSE 1-PHOSPHATE =
CC   -1- SIMILARITY: BELONGS TO THE PROKARYOTIC UDGP FAMILY.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
CC   EMBL: U03751; AAB01486.1; -
CC   DR   PFAM: PF00483; NTP_transferase; 1.
CC   DR   Transferase; Kinase; Nucleotidyltransferase.
CC   KW   SEQUENCE 279 AA; 30939 MW; D967F501 CRC32;
SQ

```

```

alignment_scores:
Quality: 96.00      Length: 236
Ratio: 0.807        Gaps: 12
Percent Similarity: 50.424      Percent Identity: 20.763

```

alignment_block:

US-09-323-427-2/rev x GALU_PSEAE ..

Align seg 1/1 to: GALU_PSEAE from: 1 to: 279

```

1559 ATTCGGTGGACAA.....GATGCGAGGTGAGCC 1528
      |||||
28  LelproValValnLysProLeuIleGlnTyrGlyVal.....Glu1 42
1527 AGAATGATGATGAGCACTTCATACAAATC..AATTTCATACAC 1481
      |||||
42  uAlaLeuAspAlaGlyLeuLysnGlnIleSerIleValThrGlyArgGlyL 59
1480 GATATGATTCGAGAGCACTTATGTAAGAGTCTTTATGATCAGCAA 1431
      |||||
59  ysaAgAlaLeuGlnAspHisPheAspIleSer.....TyrGluLeuGlu 73
1430 GGT...TGCCTATGATGATGAGGTGAGCAAGTTCGCCGAATTCAC 1384
      |||||
74  AsnGlnIleLysGlyThrAspLysGlnLysTyrLeuValGlyArgGly 90
1383 TCCATTGATTCAGCAATGTTGCCGTCACAGATCTGTGATCCAGCTG 1334
      |||||
90  sleuLeuAspLysLysSerPheSerTyrThrArgGlnThrGlnMetLysG 107
1333 GATAT.....TTTGTACACAACTGTTGTCATTTGTTGATTCACA 1293
      |||||
107  yLeuGlyHisAlaIleLeuThrGlyArgProLeuIleGlyAspGluPro 123

```

```

seq_documentation_block:
ID   CYAA_DICDI          STANDARD;          PRT:          1407 AA.
AC   003100:
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   ADENYLATE CYCLASE, AGGREGATION SPECIFIC (EC 4.6.1.1) (ATP
DE   PYROPHOSPHATE-LYASE) (ADENTYL CYCLASE).
GN   ACAA OR ACA.
OS   Dictyostellium discoideum (slime mold).
OC   Eukaryota; Dictyostellida; Dictyostellium.
CC   [1]
CC   SEQUENCE FROM N.A.
CC   MEDLINE: 92233467.
CC   PITT G.S., MILONA N., BORLEIS J., LIN K.C., REED R.R., DEVREOTES P.N.;
CC   "Structurally distinct and stage-specific adenyl cyclase genes play
CC   different roles in dictyostellium development."
CC   Cell 69:305-315(1992).
CC   -1- FUNCTION: COORDINATES CELL AGGREGATION BY SYNTHESIZING THE CAMP
CC   THAT INFLUENCES DIFFERENTIATION AND MORPHOGENESIS OF CELLS
CC   WITHIN A DEVELOPING MULTICELLULAR STRUCTURE.
CC   -1- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.
CC   -1- ENZYME REGULATION: REGULATED BY A SURFACE RECEPTOR THROUGH A
CC   GUANINE NUCLEOTIDE BINDING PROTEIN. BOTH POSITIVELY AND
CC   NEGATIVELY REGULATED BY EXTRACELLULAR CAMP; THIS REGULATION IS
CC   PART OF THE MECHANISM THAT ESTABLISHES THE OSCILLATORY CAMP WAVES
CC   DURING AGGREGATION.
CC   -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -1- DEVELOPMENTAL STAGE: CELL AGGREGATION AND LATER STAGES.
CC   -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC   -1- SIMILARITY: BELONGS TO ADENTYL CYCLASE CLASS-4/GUANYLYL CYCLASE
CC   FAMILY.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -

```



```
675 TATCAACACCCCTTGAAATTACGATGATATATCAAGCTTGCAGTTGATT 626
||:::|||||
1169 leaspthrPro.....AsnIleGluThrLysGlyIleVal 1180
625 TAGGTACCCGTGCAC 611
|||||
1181 TyrValGlnProHis 1185
```

```

1130 CCAACCGTCACAGTTCATTTGCTATCATCTGTCAGCCAGTTCATCA 1081
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
163 ProtheglyuProvalGlnPhealathrleGlynglnvalTyrHl 179
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1080 TAAATGACATGCGATTCGAAACCGTGTACTTCTGCGCGGTGTCC 1031
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179 slYstPthCysAspserGlnthrValAspPthPheCysAlaValAlh 196
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1030 ATTCCTGCTTTCGATGATGATGATACCGGTATCTGTGAATTCATAAT 981
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
196 lsserCsthrValAspAspGlyAsnGlyAspPthValGlnlleleuAsp 212
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
980 GCTGATGATGCTGCTGTGATTAATATTTGCTAATATTTGAATTC 931
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
213 GluAsnGlyCysAlaLeuAspLysPheLeuAsnAsnLeuGluTyrPr 229
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
930 AACAGATTTAATGGCTGGCCAGAGCTCAGTATACAAATATGCGATC 881
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
229 othrAspLeuMetAlaGlynglnAlaHisValTyrLysTyrAlaAspa 246
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
880 GATCACAGCTTTCTCTATCAATGCCAGATCAGTATTTACATTAAGAACCA 831
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
246 rgserrlnleuphetYrGlnCysGlnlleSerllethrValLysGluPro 262
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
830 AATAGCGAATGTGTGCACCAATGTTGAGAACCAAGATTCGGAGC 781
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
263 AsnGlnGluCysAlaArgProthrCysSerGlnProGlnleYpHeGlyAl 279
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
780 TGTTAACAGAGTGTGTCGCGCAGCAAAACCTGCTGCAGCTGCCCAATTC 731
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
279 aValLysGlnAlaAsnGlnthrAlaGln.....PhePheK 291
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
730 GTTTACTCAAGAAAGATCTGCAGAA...CCGAGATATCATTTGATGA 684
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
291 ryalLeuLysLysArgSerAlaProvalMetGlnAsnlleleuAspVal 307
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
683 CGAAGTATATCAACACCTTGAAATTTAGCATGATATCAACGTTTGGC 634
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
308 ArgAlaGlnleuthrThrleuGlnVallleuGlnGlyAsn.....LeuPr 322
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
633 AGTTGATTTTACGTCACCGTGCACCTTCTG.....CAACATATGAGC 593
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
322 oSerSerleuthrGlnAlaGlnAlaLeuValAlaSerArgGlnlleGlyG 339
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
592 AACCTGATATCTGCTGCAGTACAAATGGAATTCGATGTCACCATTT 543
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
339 luAsp.....SerPheArgGlnGlnleuGlnCyslleSerSerPhe 351
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
542 GGCTTCTCA.....ATGTTTATGGGTTTAAAGCATTCGATTCG 502
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
352 HsllleSerValValThrValPheleuGlnleuthrVal.....PheVa 366
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
501 TGCCGCTCATTTTACC.....ATTTCGTTTAAATTTTCGTCCAATTC 461
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
366 lAlaIlePheIlethrTyrMetlleValSerArgMetMetValProSerA 383
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
460 AGAG 456
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
383 sPLys 384

```

seq_name: sp_invertebrate:018479

seq_documentation_block:

```

ID 018479 PRELIMINARY; PRT; 433 AA.
AC 018479;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DE 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
DE CYPICLIN-1.
CN MRCUT-1.
OS Meloidogyne artiiella.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Diplogasteria; Tylenchida;

```

```

OC Tylenchida; Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE; 97218031.
RA DE GIORGI C., DE LUCA F., DI VITO M., LAMBERTI F.;
RT "Modulation of expression of cut-1 RNA in the
RT infective second-stage juvenile of the plant parasitic nematode
RT Meloidogyne artiiella."
RL Mol. Gen. Genet. 253:589-598(1997).
DR EMBL; X96677; CAA65452.1; -.
KW Cuticle.
SQ SEQUENCE 433 AA; 47582 MW; FDE26884 CRC32;

alignment_scores:
    Quality: 1313.50      Length: 418
    Ratio: 4.042          Gaps: 7
    Percent Similarity: 77.751      Percent Identity: 63.636

alignment_block:
US-09-323-427-2/rev x 018479 ..

Align seg 1/1 to: 018479 from: 1 to: 433

1610 ATGATTCGTTTATGCTTCTGTACTACACTTATTCATTTGCT..... 1566
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MetArgLysLeuLeuPheAlaIleGlyValPheValAlaLeuAsnAlaI 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1565 .....TATTCGATTCGGTTCGCAATGGTGCAGAGTGAGCGACG 1526
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
17 ePethrValArgAlaIleProValAspAsnGlyValGlnGlyLeuProG 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1525 AAATGAATGTGACCAACTTCATTAACAATCAATTTTAAATACGTAAT 1476
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34 lntleGluCysGlyProthrSerllethrValAsnPhenAsnThrArgAsn 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1475 GCATTCGAGGACATGTTTATGTGAAGCTTTATGATGATCAAGAGTTG 1426
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 ProPheGlnlleuHisValTyrValLysGlyLeuPheAspGlnAlaGly 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1425 CCSTAAATGATGAGGTGAGCGTCAAGTGGCGGATTTGCACTTCGATTG 1376
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 sArgSerAspGlnHisGlyArgGlnValAlaGlylleGlnLeuProPheA 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1375 ATTTCATGCAATGTTGGCGGTACAGATCTGTAATCCAGCTGATTTTGT 1326
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 sPserCysAsnValAlaArgThr...AspAlaGluProLysGlyValPhe 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1325 GTAACAACAAGCTGTTGTCATTTGCTTCATTCATCAATTTGTTACCAAGT 1276
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100 ValSerThrThrValVallleSerPheHisProGlnPheValThrLysVa 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1275 TGATGTCGATATCGAGTACAAATGCTTTTACATGGAAGCTGATAAACAG 1226
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116 lAspArgAlaTyrArgValGlnCysPheTyrMetGlnAlaAspLysThrV 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1225 TTAGTGCACAGATTGAGTATCTGAATACACAACACTGTTTCAAACTCA 1176
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
133 AlSerAlaGlnleuGlnValSerGlnlleThrThrGlnPheGlnThrGln 149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1175 ATTTCGCCGATGCGCATATGCCGTTATGAATTTGGATGGTGGACCAAC 1126
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
150 ValValPheMetProValCysLysTyrGlnlleleuGlnGlyAlaAlaLe 166
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1125 CGGTCAACCATGTTCAATTTGCTATCATTTGTCAGGCAGTTTATCAATAAT 1076
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
166 uGlyGlnProIleGlnPheAlaThrIleAlaGlnGlnValTyrThrSerG 183
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1075 GGACATGCGATTCGAAACCGTGTACTTCTGCGCGGTGTGCCATTTCC 1026
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
183 lYThrCysAspserGlnthrThrleAspThrPheCysAlaValAlhSser 199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1025 TGCTTTTCGATGATGATGATACGGGTATCTGTGAAATTTCAATGCTGA 976

```

```

RT  elegans; (OCT-1996) to the EMBL/GenBank/DBJ databases.
RL  submitted
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE: 94150718.
RA  WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.
RA  BOWFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., FULTON L.,
RA  CRATON M., DEAR S., DU Z., DUREIN R., FAVELLO A., JIER M., JOHNSTON L.,
RA  GARNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA  JONES M., KERSHAN J., KRISTEN J., LAISTER L., LATREILLE P.,
RA  LIGHTNING J., LLOYD C., MCURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA  PARSONS J., PERCY C., REFFEN L., ROOPRA A., SANDRES D., SHONKREIN R.,
RA  SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SUDSTON J.,
RA  THIERRY-MIEG J., THOMAS K., VAIDIN M., VAUGHAN K., WATERSTON R.,
RA  WATSON A., WEINSTOCK L., WILKINSON-SPOAT J., WOLDMAN P.,
RT  *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL  elegans.
RL  Nature 368:32-38(1994).
DR  EMBL: Z81088; CAB03124.1;
SQ  SEQUENCE 364 AA; 40843 MW; 3FEC30BC CRC32;

alignment_scores:
    Quality: 725.50      Length: 271
    Ratio: 3.455        Gaps: 1
    Percent Similarity: 77.491      Percent Identity: 47.601

alignment_block:
US-09-323-427-2/rev x Q9XVW7 ..

Align seg 1/1 to: Q9XVW7 from: 1 to: 364

1559 ATTCGGTGTGACCAATGCTGTGCAGAGTGAGCCAGAAATGATGTGGACC 1510
      :::::::::::::::::::: ::::: :::::
49  VALPRFLEGLASNSERLEUITYRGILYASPAVALGINALGLUCYSAPSE 65
      :::::::::::::::::::: ::::: :::::
1509 AACTTCATACATCAATTTTATACACGTAATGCAATTCGGAAGCAG 1460
      :::::::::::::::::::: ::::: :::::
65  RARGHTLLESERVALGILNILEYSTRGILYSPROPHEVALGIVALL 82
      :::::::::::::::::::: ::::: :::::
1459 TTATGTGAAGGCTTTATGATCAGCAAGGCTCGCATATGATGTAAGGN 1410
      :::::::::::::::::::: ::::: :::::
82  LEPEVALIYASPPHEALASERGLINLIVALCYSTRSERARGILYTHR 98
      :::::::::::::::::::: ::::: :::::
1409 GGAGTCACAGTGGCGGATTCCTCAATTCATTCATTCATGCAATGTTC 1360
      :::::::::::::::::::: ::::: :::::
99  GLYTRGLEUSERIALAPHELEGLINILEGLILEUICYSGILYALALE 115
      :::::::::::::::::::: ::::: :::::
1359 GCGTACACAGATCTCTGAATCCACGCTGATTTTGTAAACACACGCTTG 1310
      ::::: ::::: ::::: ::::: :::::
115  WARGILNARGVALLEUSNPRLYSGLYLEALVALARGTHRILIER 132
      :::::::::::::::::::: ::::: :::::
1309 TCATTTGCTTTCATCCATTATTTGTTACCAAGTGTGATCGTCATATCGA 1260
      :::::::::::::::::::: ::::: :::::
132  HTLLESERPHENISPROLYRPHETLHTRLYSVALASPARGTHRILYASN 148
      :::::::::::::::::::: ::::: :::::
1259 GCATACATGCTTTTACATGAGACCTGATAAACAGTTAGTCAGCAAGATGA 1210
      ::::: ::::: ::::: ::::: :::::
149  LEULLEUCYSLEUTRYARGILUSERGLINVALTHRYVALAASNAHLESE 165
      ::::: ::::: ::::: ::::: :::::
1209 GGTATCTGAATCACACACTGCTTTTCAACATCGAATGTCGCCGAGCCAG 1160
      :::::::::::::::::::: ::::: :::::
165  VALASPGIUILESERTHLLESERTYASNVALASNLEUTHMEPRROT 182
      :::::::::::::::::::: ::::: :::::
1159 TATGCCGTTATGAATTTTGGATGGTGGACCAACCGGTACACCACTGTCAA 1110
      ::::: ::::: ::::: ::::: :::::
182  HRCYSTRHYRGINLILEUSERGLYGLYPROHEGLYUPROVALGILN 198
      ::::: ::::: ::::: ::::: :::::
1109 TTTCGATATCTGTGCAGCCGATTATCATAAATGCAATGCATGCAATCTGA 1060
      ::::: ::::: ::::: ::::: :::::
199  PHEGLYLEULLEGLINGINALVALTYNHSGLITRPLYSYSPASPNAS 215
      ::::: ::::: ::::: ::::: :::::
1059 A.....ACCGTTGACTTTCTGCGCGGTGTC 1031

```



```

1  ValIleSerPheHisProLeuSerValThrIlySValAlaSParGAlaTyrAr 17
1260 AGTACAAATGCTTTTACATGAGAGCTGATAAACAGTAGTCAGCAGATTG 1211
17  gValGIncySPhetYrMeGluAlaAlaSPlySthrValSerThrGInIleG 34
1210 AGGTATCTGAATCACAACCTGCTTTTCAACATCAATGTCCTCCGATGCA 1161
34  lValIleSerGluIleThrThrAlaPheGInThrGInIleValProMetPro 50
1160 GTATGCGCTTATGAATTTTGGATGCTGGACCAACCGCTCACCATTCA 1111
51  ValCySPArTyRGIuIleLeuSpsGIyGlyProThrGIyGInProIleG 67
1110 ATTGCTATCATATTGTCAGCCAGCTTATCATCAATAGACATGCGATTTCG 1061
67  nPheAlaThrIleGIyGInProValTyRHisIlySThrPhCySPasPerg 84
1060 AAACCGGTACTTCTCTGCGCGTGTGCCATTCCTGCTTTCGATGAT 1011
84  luthrValAlaSPthrPheCysAlaValAlaHisSerCysPheValAlaSP 100
1010 GGTAAACGCTGATACTGTGGAATTTCTAAATGCTGATGATGCTGCTTGA 961
101  GlysngIySPthrValGIuIleLeuSerAlaSPGIyCysAlaLeuAs 117
960 TAAATATTTGCTAAATATTTGGAATATCCA 930
117  pLySThrLeuLeuAsnSPnLeuGIuTyRPro 127

```

seq_name: sp_invertebrate:Q21808

```

seq_documentation_block:
ID Q21808 PRELIMINARY; PRT; 290 AA.
AC Q21808;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)
DE ROT3.3 PROTEIN.
GN Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilita;
OC Rhabdilita; Rhabdilita; Rhabdilita; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA COTTAGE A.;
RL Submitted (May-1995) to the EMBL/GenBank/DBJ databases.
RX [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., FULTON L.,
RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATRILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SUSTON J.,
RA THIERRY-MIER J., THOMAS K., VAUDIN M., WATKINSON-SPROAT J., WOHLIDMAN P.;
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLIDMAN P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.
RL Nature 368:32-38(1994).
DR EMBL: Z49207; CAAB8068.1; -.
SQ SEQUENCE 290 AA; 33215 MW; 441587FE CRC32:

```

alignment_scores: 654.50 Length: 283
 Quality: 3.306 Gaps: 8
 Ratio: 69.965 Percent Identity: 45.230

alignment_block:

US-09-323-427-2/rev x Q21808 ..
 Align seg 1/1 to: Q21808 from: 1 to: 290

```

1343 AATCCACCGTGTATTTGTATACACACACTGTTGATTCCTTCATCC 1294
4  SerProArGIyLeuPheLeuSerThrAsnValAlaIlePheAsnPr 20
1293 ATTATTTGTACCAAAAGTTGATCGTCATATGAGATACAAATCTTTACA 1244
1  lValIleSerGluIleThrThrAlaPheGInThrGInIleValProMetPro 37
20  oGIuPheLeuThrIlySAsnSPArGValPheIlySValGInCysPheTyR 37
1243 TGGAGCTGATTAACACAGTTAGTCGACACAGATTGAGTATCGAATACACA 1194
37  etGIuMetGIuArGIyIleGInIlySValIleGInIleSerMetProPro 53
1193 ACTGCTTTTCAACCTCAATTTGTCGCCATGCCAGTATGCCGTTATGAAT 1144
54  ProThrMetHisSerIlySgInLeuAsnMetProValCySPlySTyRGIuVa 70
1143 TTTGGATGTGGACCAACCGTCAACAGCTTCAATTTGCTATCATGTC 1094
70  lLeuAspGIySerProThrGIyProProValTyRPhAlaThrValGIyG 87
1093 AGCCAGTTTATCATTAATGGACATGCGATTCTGAACCGTTGATACCTTC 1044
87  lMetValTyRHisIlySThrPhCySPthrGIuHisGIuAsnThrPhe 103
1043 TGCGCGGTGTTCATTCCTGCTTTCGATGATGATGATGATGATGATGATGAT 994
104  CysMetLeuValHisSerCysPheValAlaSPspGIyAsnGIyGInArGIy 120
993 GGAATTTCTAAATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 944
120  lGInLeuLeuAsnSPlySGIyCysAlaLeuAspIlySTyRLeuLeuThr 137
943 ATTGGAATATTCACACAGATTAAATGCTGCGCAAGAACCTCAGTATTC 894
137  snLeuGIuTyRProThrAsnSPleuMetAlaGIyArGIuIleHisValTyR 153
893 AATATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 844
154  lYsTyRAlaAlaSPArGIySPAsnMetTyRPhAspCySGInIleSerIleTh 170
843 CATTAAAGAACCAATATGCGAA...TGCTGTCAGCACCATGCTTCGAGAC 797
170  rValIlySGluProGIyLeuAspTyRGIySAspValProSerCySPAsp 187
796 CA..... 795
187  roProArGIyArGIySerAsnThrLeuProAlaProAspAsnIle 203
794 .....CANGATTCGAGACTGTTAAAC 772
204  ThrAlaIleAlaIleHisIleGIuTyRGIuAspSerGIuIleIleSerA 220
771 AG...GTGTCGCCACAGCAAAACCTGCTGACGTC.....CGC 737
220  sPTyRlIleIleProAsnSPAspIleIleSerIleAsnThrIleGIuArGIy 236
736 AACTTCGTTTACTCAGAAAGATCTGCAGAACCGGAGAGATATTCATGAT 687
237  AsnPhe.....AspMetArGIyIleSer..... 243
686 GTACGAAGCTGATATCAACACCTTGAAATTAAGCATGATATCAAGCTTT 637
244  ...GIuLeuCySMetThrAlaIleGIyThrThrLeuLeuValPheIleVa 259
636 GCCAGTTGATTACGTC.....ACCGTGCATCTTTCG 605
259  snAlaPheLeuPheIleIleSerLeuValSerIleValAlaHisValCyS 274

```

seq_name: sp_invertebrate:Q21540

RP SEQUENCE FROM N.A.
 RA MCMURRAY A.;
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
 RN 121
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 DR EMBL; 266496; CAA91280.1;
 SQ SEQUENCE 484 AA; 53325 MW; 6AEFC669 CRC32;

alignment_scores:
 Quality: 595.00 Length: 251
 Ratio: 3.199 Gaps: 4
 Percent Similarity: 74.104 Percent Identity: 45.020

alignment_block:

US-09-323-427-2/rev x Q19053 ..

Align seg 1/1 to: Q19053 from: 1 to: 484

1535 GGTGAGCCGAATGATGATGCGACCACTCAATACATCAATTTAA 1486
 |||||
 14 GlyGluProGluValValcysGluThrAlaSerIleSerIleuLeu 30
 1485 TACAGCTAATGATGATGAGCATGTTATGGAAGCTTTATGATC 1436
 |||||
 30 sThrArgAsnSerPheAsnGlyValIlePheValLysGlyTyrValSerG 47
 1435 AAGAAGTGGCGTAATGATGATGAGCGCAAGTGGCGGAATTTCA 1386
 |||||
 47 InProSerCysMetThrValGlyAspLysThrGlyHisArgPheGlu 63
 1385 CTTCATTTATTCATGCAATGTTGGCGTACAGATCTGTGATCCACG 1336
 |||||
 64 ValArgHisAspSerCysGlyValArgGlnArgGlnIleAsn.... 78
 1335 TGGTATTTTGTATACACAACTGTTGATTCCTTCATCATCATTTTG 1286
 |||||
 79 GlyValValIleSerAlaThrValIleIleSerPheHisSerIlePheI 95
 1285 TTACCAAGTGTGATGCGTATGATGAGTACAGTCTTTTACATGGA 1236
 |||||
 95 LeuHisValAspArgAlaThrValIleIleSerPheHisSerIlePheI 111
 1235 GATAAAGAGTTAGTGCACAGATTGAGGTATGCAATACACACGCTT 1186
 |||||
 112 ThrLysLysValHisAsnHisValAspIleSerAlaLeuThrGlnI 128
 1185 TCMAAATCAATTTGCCGATGCCGATGCGGTATGCAATTTTGATG 1136
 |||||
 128 uLeuGluSerGluThrGlnLeuProValCysArgTyrGluIleLeuAs 145
 1135 GTGGACCAACGGGTCAACAGTTCAATTTGCTATCATGTCAGCCAG 1086
 |||||
 145 IuAla...GlyGlySerProIleLysTyrAlaArgIleGlyAspGln 160
 1085 TATCATTAATGATGATGATGATGATGATGATGATGATGATGATG 1036
 |||||
 161 TyrHisLysTyrPheCysValAlaGluLeuGluAsnValTyrCys 177

1035 TGTCCATTCCTGCTTTGTCATGATGCT...AACGTGATACTGTGAAA 989
 |||||
 177 sValHisSerCysThrValTyrAspGlyGlnGlyGlyProValThrV 194
 988 TTCTAATGCTGATGATGATGATGATGATGATGATGATGATGATG 939
 |||||
 194 AlIleAspAlaAsnGlyCysSerValAspGlyValIleLeuGlnAsn 210
 938 GATATTCACACAGATTAATGCTGCGCCAGAGCTCACGATACAAAT 889
 |||||
 211 GluTyrThrSerAspLeuThrAlaGlyLysLeuAlaProValPheLys 227
 888 TGGGATGATGATGATGATGATGATGATGATGATGATGATGATG 839
 |||||
 227 eAlaAspLysAlaGlyLeuTyrPheAsnGlyGlnIleGlnLeuThr 244
 838 AAGACCAATATAGGATGATGATGATGATGATGATGATGATGATG 795
 |||||
 244 yAspValAsnTyrGlyCysSerAsnThrGlnProGlnCysProThr 260
 794 CAA 792
 |||||
 261 Gln 261

seq_name: sp_invertebrate:Q93532

seq_documentation_block:
 ID Q93532 PRELIMINARY; PRT; 315 AA.

AC Q93532;
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
 DE F20D1.8 PROTEIN.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;

OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

RN [1]

RP SEQUENCE FROM N.A.

RA BURTON J.;

RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE; 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,

RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,

RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,

RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,

RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,

RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,

RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,

RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,

RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,

RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLMAN P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.";

RL Nature 368:32-38(1994).;

DR EMBL; 278542; CAB01742.1;

SQ SEQUENCE 315 AA; 35785 MW; CD2B4B3D CRC32;

alignment_scores:
 Quality: 580.50 Length: 292
 Ratio: 2.888 Gaps: 6
 Percent Similarity: 68.836 Percent Identity: 39.384

alignment_block:

US-09-323-427-2/rev x Q93532 ..

Align seg 1/1 to: Q93532 from: 1 to: 315

1622 TCTTCTGAATGATGATGATGATGATGATGATGATGATGATGAT 1584
 |||||

```

14 ThrSerLysAlaIleValHisIleIlePhePheTrpTyrLeuGlyAsnAr 30
1583 ....ACACTTATGCAATGCTTATTCATCCGCTGACAAATGGTTCG 1538
30 GILEProIleuPheCysLysTyrHisIleGluGlnIleAspAnGlyLeuG 47
1537 AAGTGACCCGAAATGAAATGATGGACCAACTGCAATTAACATCAATTT 1488
47 InGlyGluProIleuIleArgCysGlySerGluSerLeuSerIleAsnPe 63
1487 AATACACCTAATGCAATTCGAAAGACATGTTATGTAAGAGCTTTATGA 1438
64 LysThrGlnGlyAlaPheGluGlyHisValTyrValLysGlyHisTyrSe 80
1437 TCAGAAAGCTGGCCGATGATGAAAGTGGACGCAATGCGCGAATTT 1388
80 rMetLysHisCysArGlnTrpAlaThrLeuGlnIleSerGlnValAsnLeu 97
1387 CACTTCATTTGATTCATGCAATGTTGGCGGTACAGATCTCTGATGCA 1338
97 hrValSerTyrSerAlaCysAspValIleArgGlnArgSerSerAsnPro 113
1337 CGGTGATTTTGTAAACAACACTGTTGCTTTCGTTTCATTCATTAAT 1288
114 LysGlyIleMetMetThrAlaThrIleIleIleSerPheHisPrometPh 130
1287 TGTATACAAAGTATGATCGATATCGAGTATGCAATGCTTACATGAG 1238
130 eIleThrLysIleAspLysSerTyrLysValGlnCysPheTyrAlaGlu 147
1237 CTGATAAACAGTATGAGTACAGATGAGTATCT..... 1203
147 IagLysLysThrValThrGlnGlnIleuAsnValAspIleAlaLysGluGln 163
1202 .....GA 1201
164 GluLysLysIlePheValMetValGlyAspGluGlyGlyThrValSe 180
1200 AATCACAATGCTTTTCAACCAATTT..... 1173
180 rHisThrThrGlyAspGlnLysLysLeuHisLysLeuAsnAspProSerT 197
1172 .....GTCCGATGCCAGTATGCGCTTATGAA 1146
197 hrGluGluArgIleSerTyrAsnValProIleuProAspCysLysTyrArg 213
1145 AATTGGATGGTGACCAACCGGTCAACAGTCAATTTGCTATCATGG 1096
214 ValLeuThrGluSerLysThr...GluGluValAlaPheAlaThrValG 229
1095 TCAGCCAGTTTATCATTAATGAGATCGATTC.....GAACCG 1055
229 yGlnIleValTyrHisGluTrpSerCysGluAlaProGlyGlnAsnLnt 246
1054 TTGATATCTTTCGCGGTTGTCATTCCTGTTTGTGATGATGATAAC 1005
246 hrSerProPheCysValThrValHisSerCysAsnValLysAspGluThr 262
1004 GGGATACGTGTGAATTTCAATGCGTGAATGATGCTCTTGATAATA 955
263 GLyLysGluValGlnIlePheAspGluAsnGlyCysAlaValAspLysTy 279
954 TTTTGCTAATTAATTTGGAATATCCACAGATTAATGCGGCGCAGAG 905
279 rLeuIleAsnAsnLeuGluTyrSerSerAspLeuThrGlyGlnLeuS 296
904 CTCACGTATTAACAATATGCGGATGCA 879
296 eArgLysValCysSerTyrThrValArg 304
seq_name: sp_invertebrate:Q22164
seq_documentation_block:

```

```

ID      Q22164      PRELIMINARY;      PRT;      384 AA.
AC      Q22164;
DT      01-NOV-1996 (Tremblrel. 01, Created)
DT      01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT      01-JAN-1999 (Tremblrel. 09, Last annotation update)
DE      T04F8.4 PROTEIN.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC      Rhabditina; Rhabditoidea; Rhabditidae; Peloderiinae; Caenorhabditis.
RN      [1]
RP      SOURCE: FROM N.A.
RA      LENNARD N.;
RL      Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE: 94150718.
RA      WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA      BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA      CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA      GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA      JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA      LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA      PARSONS J., PERCY C., RIKKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA      SMALDON N., SMITH A., SONNHAMMER E., STADEN K., SUSTON J.,
RA      THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA      WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOLDMAN P.;
RT      "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT      elegans."
DR      Nature 368:32-38(1994).
EMBL: Z66565: CA91480.1; -.
SQ      SEQUENCE 384 AA; 43964 MW; 9FFEF44 CRC32;

```

```

alignment_scores:
Quality: 546 50      Length: 394
Ratio: 2.240      Gaps: 9
Percent Similarity: 61.929      Percent Identity: 31.218

```

alignment_block:

US-09-323-427-2/rev x Q22164 ..

Align seg 1/1 to: Q22164 from: 1 to: 384

```

1601 CTTATTCCTTCTGTACTACATTAATGCAATGCTTATTCATGATCCGCT 1552
10 LeuValGluPheThrTyrPheValLeuGlyPheSerAlaIleGlnAs 26
1551 TGCAATGAGTGTGCAAGGTGACCGCAAGAAATTAATGTGACCAACTGAA 1502
26 rAspAsnGluLeuIleGlyGlnProGluIleGlnCysAsnAlaAspThrI 43
1501 TAACAAATCAATTTTAATACAGTAAATGCAATTCGAAAGACATGTTATG 1452
43 LeAspMetGlnPheAlaGlyThrArgLysGlnPheAsnGlyLysValTyrVal 59
1451 AAAGCTTTATGATCAAGAAGCTTCCGT.....AATGA 1417
60 LysGlySerTyrAsnArgProGluCysArgValAspTyrSerThrLysAs 76
1416 TGAAGTGGACGTCGCAAGTTCGCGAATTTCACTTCATTTGATTCATGCA 1367
76 rGlnPheGlyArgProValGlyGlyIleLysLeuAsnHisGlyAlaCysA 93
1366 ATGTGCGCTACAGCATGCTGTGATTCGACAGTGTATTTTGTGAACAACA 1317
93 smetAspArgGlnArgMetIleAlaProGluGlyMetCysPheSerThr 109
1316 ACTGTTGCAATTTGTTTCAATCAATTTGTTTCAATTTGTTTCAATTTGAT 1267
110 ValLeuIleIleSerPheHisProLeuPheLeuThrArgMetCysPrysa 126
1266 ATATCGATTAACAATGCTTTTACATGGAAGCTGATAAACAAGTTATGCA 1217

```

```

|||||.....|||||.....|||||.....
126 atyrhisilearqysmetrrylgsluhalaharvnlrvala 143
1216 AATATGAGTATCTGAATACACATGCTTTTCAACCTCAATGTCGG 1167
|||||.....|||||.....
143 lallspalserasnleuprthrcluservalglinseraspleupro 159
1166 ATGCCAGTATGCCGTATGAAATTTTGATGAGTGACCAACCGTCAC 1117
|||||.....|||||.....
160 MetPrthrCysSerYrThrIleArgArgAspGlnLeuAspGlyProII 176
1116 AGTTCATTTGCTATCATCTGTCAGCCAGTTTATCATTAATGACATCG 1067
|||||.....|||||.....
176 eleuylstyralalysvalgllyasglnvalhlsargtrpGlnCysA 193
1066 ATTTGAAACCGTTGATACTTTCGCGCGGTTCCTCATTCCTGTTGTC 1017
|||||.....|||||.....
193 spserglu.....AspYrclyleuLeuValHlsSerCysYrVal 206
1016 GATGATGTAACGGTATCTGCTGGAATTTCTAATGCTGATGATGTC 967
|||||.....|||||.....
207 GlusprgylngllyglulysglmetlilellespgluatrglyCysHl 223
966 TCTTGATTAATTTGTAATATTTGAAATATCCAAACAGATTTA...A 920
|||||.....|||||.....
223 sthrAsprglleuLeuLenglyAspProthrTyrValGlulalAsm 240
919 TGGCGGCCAAGAACCTCAGTATACAAATATGCGGATGATCAGACGCT 870
|||||.....|||||.....
240 etalTyrArgIserPheValPheIAsprgllealaval 256
869 TTCTATCAATGCCAGATCAGTATACCAATTAAGAACCAATAGCGAATG 820
|||||.....|||||.....
257 ArgPheGlnCysGlnleargLeuLysAsp...AspGlyGlyCys 272
819 T.....GTTGACCAATGTTGAGAACCCACAGATTCGAGCTG 779
|||||.....|||||.....
272 saspglyIlethrProPrometCysSer.....P 282
778 TTTAAACAGGTGTCGCCGACAAACCTGCTCAGCTGCGCAACTTCGT 729
|||||.....|||||.....
282 helysAspAlaAsnSerAsnGlnleValIlysalArgAsnAlaArg 238
728 TTATCTACAGAAAGATCTGACGACCGAGATATCATGATGATGACAC 679
|||||.....|||||.....
299 ThrPheLysProLysAspAlaAsp.....MetPheSerGlnThrVa 312
678 TGAATATCAACACCTTGAAATTTAGCATGATATCAGCTTGGCCACTG 629
|||||.....|||||.....
312 lTyrValMetAspLysGlnAsnGlyAspSerThrSerAlaGlnAlaIle 329
628 ATTACGTCACCGTCGACCTTCTGCAACATATGACAAACCTGTAATCT 579
|||||.....|||||.....
329 luleIArgGlu..... 332
578 GCTGCAGTACAAATGAAATCTGCATGCACCA.....TTGG 541
|||||.....|||||.....
333 ..LeuAspProGlnThrIleCysLeuAlaProLysLeuLeuValAla 348
540 CTTCATGATGTTATGAGTTTAAAGCATGATGATGCTGCCGCGCATTA 491
|||||.....|||||.....
348 aValThrPhePheThrPheIlePheValLeuPheValThrThrIleLeu 365
490 TTACCATTTGCTTTAAATTTGCTCCAAATCAG 459
|||||.....|||||.....
365 alValValIlyThrHisArgTyrCysLysAsnGlu 375
seq_name: sp_invertebrate:Q9XVNZ
seq_documentation_block:
ID Q9XVNZ PRELIMINARY: PRT: 610 AA.
AC Q9XVNZ2
DT 01-NOV-1999 (Tremblrel. 12, created)

```

```

DT 01-NOV-1999 (Tremblrel. 12, last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, last annotation update)
DE F53B6.6 PROTEIN.
GN F53B6.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WHITE S.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RN submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
RP [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSKOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., KERSHAW J., HAWKINS T., HILLER L., JER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., PERCY C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA SMALDON N., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA THIERRY-MIEG J., THOMAS K., WILKINSON-SPROAT J., WOHLDMAN P.;
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL: Z81086; CAB03116.1;
SQ SEQUENCE 610 AA: 68585 MW: D793181D CRC32:

```

```

alignment_scores:
  Quality: 513.50      Length: 298
  Ratio: 2.593        Gaps: 8
  Percent Similarity: 66.443      Percent Identity: 35.906

```

alignment_block:

US-09-323-427-2/rev x Q9XVNZ ..

Align seg 1/1 to: Q9XVNZ from: 1 to: 610

```

1652 TTAGTCTTCTTAAATCGAATTACTTAATCTTGAATGATGATTCG 1603
|||||.....|||||.....
34 leuArgPheGlnSerValGlnPhe...LysAsnProIleLeuLeu 49
1602 TCTTAT.....GCTTCTGTACTACACTTATTCAT 1571
|||||.....|||||.....
49 uleuIleYrAsnArgIlePheLeuGlnSerIleThrIleLeuLeu 66
1570 TGTCTTATTCG.....ATTCCGGTTGACAAATGCTGCGAAGTGACCA 1527
|||||.....|||||.....
66 euserThrSerCysPheGlnleuGlnAsnGlyValAlaGlyLysPro 82
1526 GAAATTTGAATGAGTGCACCACTTCAATATACAAATTTTAATACAGTAA 1477
|||||.....|||||.....
83 GluValPheCysGlyIleAspThrIleArgValIlyValAsnThrGlnHl 99
1476 TGCATTCGACAGACATGTTTATGAAAGTCTTTATGATCAAGAAGTT 1427
|||||.....|||||.....
99 sProPheAsnGlnArgIleTyrValAspGlyLysSerAspLysGlnHSC 116
1426 GC.....CGTAATGATGAAGTGGACGTCGAAGTTCCCGGAATTTTCACAT 1383
|||||.....|||||.....
116 yValAlaGlnHisSerAlaAspAlaHisSerSerProGlnIlyPheThrIle 132
1382 CCATTTGATTCATGCAATGCTGCGCGCTACACGATCTCTGATTCACAGTGG 1333
|||||.....|||||.....
133 proIleGlyAlaCysAsnMetArgGlnArgThrLeuHisProArgGln 149
1332 TATTTTGTAAACAACAAGTGTGTCAATTCGTTTCATTCATTAATTTGTTA 1283

```



```

1111
149 yllserperphenmetilserphenhisprophepvalt 166
1282 CCAAGTGGATGCGATATCGACATGCTTTTACATGGAAGTGT 1233
166 hglymeaspargalapheSerileargCyspheleuGlusterile 182
1232 AAACAGTTCAGTCACAGATTGAGTATCTGAATCAGACAGCTTTTCA 1183
183 LysglYleuAsnAlaGluileaspValGlythrLeuAlaProGlnHisVa 199
1182 AACTCAAAATTCGCCGATGCCGATATGAAATTTTGATGTG 1133
199 LaspInGluYrSerLeuProValCysAlaTyrHisLeuLysasp...G 215
1132 GACCAACCGGTCACACAGTTCATTCATTCATTCGTCAGCCAGTTTAT 1083
215 LylleGluGlnHisValleuArgPheAlaGlnValGlnLysValThr 231
1082 CATAAATGAGATCGATTCGAAACCGTTCATCTTCTGCGCGTGT 1033
232 HisValItrparGysAspGlnAspAlaSerHisValItrGlyIleLeuIl 248
1032 CCATTCCTGCTTTGTGATGATGTAACGATGATCTGGAATTTCTAA 983
248 ehIsSerCysTyrAlaAspAspGlnHisGlnLysPheGlnLeuValA 265
982 ATGCTGATGATGCTGCTGTGATAATTTGCTTAATTTGGAATAT 933
265 spAspArgGlyCysSerThrAspProPheLeuProGlnIleGluYr 281
932 CCAACAGAT...TTAATGCTGCGCCAGACAGTCACGATACAAATATGC 886
282 GluHisGlnAlaIleSerAlaItrThrAsnAlaHisValIlePheYrAl 298
885 GGATGATGATGATGCTTTCTATTCATGATGCCAGATGATTTACATTAAG 836
298 aAspLysValGlnLeuYrPheThrCysThrValGlnLeuCysTyrLysH 315
835 AACCAATAGGGAATGT.....GTTCACCAACATGTTCA 801
315 Is...AspGlyGlyCysGlnGlyIleThrProProGlnCysSer 328
seq_name: sp_invertebrate:Q23097

seq_documentation_block:
ID 023097 PRELIMINARY: PRT: 387 AA.
AC 023097:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DE W01A8.3 PROTEIN.
GN W01A8.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditiida;
OC Rhabditiina; Rhabditiodea; Rhabditiidae; Peloderinae; Caenorhabditiis.
RN [1]
RP SEQUENCE FROM N.A.
RA WILKINSON J.:
RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BOWFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAMKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KESHAW J., KIRSTEN J., LAISTER N., LATEBRIE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., REIKEN L., ROOPRA A., SAUNDERS D., SHOMKNEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SOLTSON J.,
RA THERRER-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
```

```

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
DR EMBL; 271267; CAA95850.1;
SO SEQUENCE 387 AA; 43212 MW; D127FF23 CRC32;

alignment_scores:
Quality: 479.00 Length: 383
Ratio: 2.10 Gaps: 11
Percent Similarity: 59.269 Percent Identity: 31.593

alignment_block:
US-09-323-427-2/rev x Q23097 ..

Align seg 1/1 to: Q23097 from: 1 to: 387

1529 CCAAGAAATTCGAATGTGGACCAACTTCATTAACATCAATTTTAATACAG 1480
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
48 ProLysValIleCysAlaGlnAsnAspLeuAlaLeuAspIleValThrSe 64
1479 TAAATCATTCGACAGACATGTTTATGTGAAGGCTTTATGATCAAGANG 1430
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
64 rlySproheargGlyAsnIlePheValLysGlyArgAlaLysAspLysS 81
1429 GTTGCCGTAATGATGACAGTGCAGCTCAAGTTCGCCGATTTTCACCTTCA 1380
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
81 erCysArgGlnSerTyrAlaAsnAsnGlyThrAsnSerTyrSerLeuPro 97
1379 TTTGATTCATCATGTTTGGCCGTCACAGATCTCTGAATCCACGTGTAT 1330
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98 LeuGlyLysCysGlyMetGlnArgLeuArgSerAlaAsnProArgGlyLys 114
1329 TTTTGTAAACAACAACGTGTTCATTCATTCATTCATTCATTCATTCATTC 1283
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
114 IAsnPheMetValThrValIleValSerPheHisProAlaGlyPheIleT 131
1282 CCAAGTTCGATGTCGATATGATGATGATGATGATGATGATGATGATGAT 1233
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
131 hrlYsAsnAspArgAlaPheHisValLysCysPheYrMetGlnProAsp 147
1232 AAACAGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1183
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
148 GluIleValThrGlnAsnIleAspValSerMetIleProThrThrGluLe 164
1182 AACTCAAAATTCGCCGATGCCAGATGCGGATGATGAATTTTGGATGTG 1133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
164 userAspSerMetValMetProLysCysGluYrSerValArgArgAspG 181
1132 GACCAACCGGTCACACAGTTCATTCATTCATTCATTCATTCATTCATTCAT 1083
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 LyrProAsnGlyProThrLeuThrTyrAlaAsnValGlnAspIleValPhe 197
1082 CATTAATGACATGCGATTCGTAACCGTTCGATTCATTCATTCATTCATTCAT 1033
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
198 HisValItrpLysCys.....ThrProAlaAspMetGlnLysLeuVal 211
1032 CCATTCCTGCTTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 983
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
211 LysLysCysPheValThrAspArgLysAspGlnLysAspHisAlaValAla 228
982 ATGCTGATGATGCTGCTTCGTAATATTCGTAATATTCGTAATATTCGTAAT 933
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
228 spPheAspArgLysAlaItrAspProPheLeuLeuSerGlnLeuSerTyr 244
932 CCAACAGATTTAATG...GCTGCCAAGAAAGCTACCGTATACAAATATGCG 886
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
245 AspAlaSerLeuMetArgAlaHisAlaSerSerGlnValIlePheYrAl 261
885 GGATGATGATGATGCTTTCTATTCATGATGCGGATGATGATGATGATGATGAT 836
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
261 aAspSerAsnGlnLeuYrPheThrCysGlnIleArgLeuCysGlnLysG 278
```

```

835 AACCAAT.....ACGATGTGTGACACATGT..... 804
      :::::|||||:|||||:
278 lmetgymetCysgIngluValThrProAsnGlyValLys 294
803 .....TCAGACCCAGAGATTGAGCTGTTAAACAGGTGTCGCC 760
      |||||:|||||:
295 LeuLeuSerGluAlaSerGlyAspGly..... 303
759 AGCAAAACCTGCTGACGTCGGCACTTCCTTACTCAGAAAGATCTG 710
      :::::|||||:
304 .....AsnArgTyrIle 307
709 CAGAACCGGAGATATCATGTATGATGACACTGATATCAACACCTTGAA 660
      :::::|||||:
307 ysArgGluAlaAspArgSerAspTyrGluIleAspValAlaIleThrSerIle 323
659 ATTAGCGATGATAATCAAGCTTGTGCCAGTGTATTCAGTCCGTCGACT 610
      :::::|||||:
324 Leu.....LeuValLeuAspProAlaAspArgGlyLe 334
609 TCTGCACATAATGAGACAACTGTATACTTGTCTGACGACAAATGGA 560
      |||||
334 uLeu..... 335
559 TCTGCATGTCAACATTTGGCTTCTCAATGTTATG.....GGTTTACG 516
      :::::|||||:
336 ..AlaProSerProPheCysValProArgLeuLeuLeuProValLeuPro 351
515 ATTGCATGATGTCGCGCTCATTTATACCATTTGCTTAATTTGCTGC 466
      :::::|||||:
352 LeuIleLeuIleThrIleValSerLeuThrValVal.SerThrAlaLeu 368
465 AATTCAGAGG.....CATAAATAATGTTAGATTCATCGACGA 425
      |||||
368 alleArgArgGlnAsnHisLysGluLeuAspIleMetGlnSer 383
seq_name: sp_invertebrate:Q93967

seq_documentation_block:
ID Q93967 PRELIMINARY; PRT: 440 AA.
AC Q93967; Q94405;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)
DT 01-JAN-1999 (TREMBLrel. 09, last annotation update)
DE ZK265.8.
GN ZK265.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderiinae; Caenorhabditis.
OC
RN SEQUENCE FROM N.A.
RP DOBSON R.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BINFELD J., BURTON J., CONNELL M., CORSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAYELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THERRY-MEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOLDMAN P.; III of C.
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN
RP SEQUENCE FROM N.A.
RL LENNARD N.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.

```

```

DR EMBL; 281143; CAB03520.1; -.
DR EMBL; 275713; CAB03520.1; JOINED.
DR EMBL; 275713; CAB00057.1; -.
DR EMBL; 281143; CAB00057.1; JOINED.
SQ SEQUENCE 440 AA; 49299 MW; 3F2BBBA2 CRC32;

```

```

alignment_scores:
  Quality: 477.50      Length: 320
  Ratio: 2.296        Gaps: 9
  Percent Similarity: 65.000      Percent Identity: 33.438

```

```

alignment_block:
US-09-323-427-2/rev x Q93967

```

```

Align seg 1/1 to: Q93967 from: 1 to: 440

```

```

1607 ATTCGCTTATGCTTCTTGTGACTACACTTATTCATGTTGCTTATTCAT 1558
      :::::|||||:
4 LeuIleIleLeuIleIleAlaIleThrGlyIleThrArgValAlaAsnAlaIle 20
1557 TCCGCTTACATGCTGTGCAAGTGCAGCCAGAAATGAAATGTGACCAA 1508
      |||||
20 eSerIleAspAsnGluIleIleGlyGluProAspIleGlyCysLeuGlu 37
1507 CTTCATTAACAATCAATTTTATACACGTAATGCAATTCGAGACATGTT 1458
      :::::|||||:
37 spGluIleArgIleIleThrValLysThrArgLysIlePheAlaGlyArgIle 53
1457 TATGTGAAGGCTTTATGATCAAGAGGTGCCGTAATGATGAAGAGT 1408
      |||||
54 TyrAlaIleGlyArgAlaGluLeuGluAspCysTyrLysAspPheIle 70
1407 AGCTCAAGTTGCC.....GGAATTCACCTTCATTTGATTCATGCA 1367
      :::::|||||:
70 ysAsnGluLysThrArgLysProHisPheAspLeuGlnPheGlyAlaCysG 87
1366 ATGTCCGCGCTACACGATCTGTGATCCAGCTGATTTTGTGAACACA 1317
      :::::|||||:
87 LysMetLysSerLeuArgSerValAspProArgLysMetYrGlyIle 103
1316 ACTGTGTCAATTCGTTTCATCCATTTATTCACCAAGTTGATGCTGC 1267
      |||||
104 ThrValValAlaSerPheHisProLeuPheIleThrLysAlaAspGlnAl 120
1266 ATATCAGATCAATGCTTTATCATGAGCTGATTAACAGTAAAGTATGTCAC 1217
      |||||
120 atYrHisValLysCysPhePheGluGluAlaAsnLysGlyLeuThrAlaG 137
1216 AGATTGAGGTATCTGAATACACAACAGCTTTTCAACTCAATGATGTCGG 1167
      :::::|||||:
137 IuLeuGlyValSerMetIleProThrThrGluLeuGluAlaArgHisGly 153
1166 ATGCCAGTATCCGTTATGAAT.....TTGA 1138
      :::::|||||:
154 IleProGlyCysThrTyrSerIleHisArgSerThrIleAspGluLeuAs 170
1137 TGTGTGA...CCAACGGGCAACAGTCAATTCCTATTCATGTCGACG 1091
      :::::|||||:
170 PAlaGlyArgProAlaGlyAsnValIleGlnPheAlaArgValGlyGlu 187
1090 CAGTTTATCAATATGACAGATGCGATTGCAACCGTGTGATCTTTCGCG 1041
      |||||
187 rValLeuHisGlnIleThrPheLysAsnAspGln.....MetYrGly 200
1040 GCGGTGTTCATTCCTGCTTGTGATGATGTAACGGTATGATGTCGGA 991
      :::::|||||:
201 ValLeuIleAsnAsnCysTyrValAlaThrAspGlyPheGlyLysAlaAs 217
990 AATTTCAATGCGATGATGCTGCTGTGATAAATATTGCTAATAATAT 941
      :::::|||||:
217 pValIleAspAspLysGlyCysProIleAspProIleLeuIleThrGlyI 234

```

factor: phenylalanyl-tRNA synthetase; RNA binding.
 Caenorhabditis elegans.
 ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae; Rhabditina; Rhabditoidea; Rhabditidae; Pelodierinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 25823)
 Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, D., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durlin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, U., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, R., Percy, C., Rifkin, L., Roopra, A., Saudek, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkerson, P., and Wohldman, P.
 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
 Nature 368 (6466), 32-38 (1994)
 JOURNAL MEDLINE 94150718
 REFERENCE 2 (bases 1 to 25823)
 Sims, M.
 TITLE Direct Submission
 JOURNAL Submitted (13-JUL-1995) Louis, MO 63110, USA. E-mail: jess@sanger.ac.uk or rwenemate@wustl.edu
 COMMENT Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information.
 For a graphical representation of this sequence and its analysis see: -
 http://webace.sanger.ac.uk/cgi-bin/displayJob-wormaccetclass-SequenceJobect-F22B5
 Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
 IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring subclones.
 IMPORTANT: This sequence is not the entire sequence overlapping sections. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring subclones.
 The true left end of clone F22B5 is at 24607 in sequence 266522.
 The true right end of clone F22B5 is at 16002 in sequence 266523.
 The true left end of clone M05D6 is at 25723 in this sequence. The true right end of clone F14B5 is at 4610 in this sequence. The start of this sequence (1..101) overlaps with the end of sequence 266522.
 The end of this sequence (25723..25823) overlaps with the start of sequence 266523.
 FEATURES
 source Location/Qualifiers
 1..25823
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /chromosome="II"
 /clone="F22B5"
 411..1471
 /gene="F22B5.2"
 join(411..462,584..663,745..1127,1216..1471)
 /gene="F22B5.2"
 /note="similar to RNA binding protein; cDNA EST yk390d8.5 comes from this gene; cDNA EST yk499d6.3 comes from this gene"
 gene
 /codon_start=1
 /protein_id="CAA90354.1"
 /db_xref="GI:3876226"
 /db_xref="SWISS-PROT:Q19706"

/translation="MAPAEVVSNAEAEVDQNAHIOGADGTRTETAFTEVDGVRNK
 VTQFQVINKRVKVVADRRKMKVFGSGCGEPAGQVATITVAEEVMOFTRNAGEO
 ILVQEDKQTAKTTSREHCRKCGNDHSTGCPYKVMQDEADADKDTKDMNMG
 MRPGROITDRNSDENTCVRNTLPOEMEDLRLDFGKIGVIRIFIARDKVTGLPKG
 FAETVFESRDDAARAIALINDIRNHWLVKAEWTRPSN"
 1870..2668
 /gene="F22B5.1"
 join(1870..2158,2209..2387,2534..2668)
 /note="similar to GTP-binding ADP-ribosylation factor"
 /codon_start=1
 /protein_id="CAA90353.1"
 /db_xref="GI:3876225"
 /db_xref="SWISS-PROT:Q19705"
 /translation="MGFLIKRKAREREMKILILGLDNACKTLLMKFDEPDPTI
 EPTLPDPTKTVFHPDQLNMDYGGKSLRSYKNVFESTDALIWDSSDDERLDQ
 SEELKLQERLQAGSILLVANKSDLPALIVNSIAGVITQMSVYTSSEKCFVLIDL
 HSLSHHWKIFSCALSGRLVQATWLCDDVDGSRFLFD"
 complement(2833..4393)
 /gene="F22B5.3"
 complement(2833..3042,3088..3700,3823..4048,
 4273..4393)
 /gene="F22B5.3"
 /note="similar to cuticulin; cDNA EST EMBL:T01970 comes from this gene; cDNA EST CEBSX90RB comes from this gene; cDNA EST EMBL:T00412 comes from this gene; cDNA EST yk386c7.3 comes from this gene; cDNA EST yk386c7.5 comes from this gene; cDNA EST yk504b12.3 comes from this gene; cDNA EST yk543g2.3 comes from this gene; cDNA EST yk652b8.3 comes from this gene"
 /codon_start=1
 /protein_id="CAA90355.1"
 /db_xref="GI:3876227"
 /db_xref="SPRMBL:Q19707"
 /translation="MARYSLGICGLIVASVAPVDNNEGEPEVCGTSTVNFN
 TRNFESEHVYKGLFDQDQCRNDGSGVAGIELPDTQCNVARRSLNPGVETTV
 VVSHFQPVYKVDRAVVOCHYKMDADTVSQITVSPQLTAFOGVMPICKYEILN
 DGFLNPEVQFATIQOYVHKWTCDSFVDFCAVHSGCTDDGSDVQVILDEGCL
 KGPILLNEXPTDLMACQEAHVYKADRSQLEFYCQSTIVKEPNECARPTCGEPG
 FGAVKQNOTQOFYRLKRSAPVMENTLDVRAELTLEVEGLNPSLTOAQLVAS
 REIEDSDRQELCSSPHISVYVFLGIVAFITVIVSRMWSPSKQMSAC"
 9691..10234
 /gene="F22B5.4"
 join(9691..9890,9940..10234)
 /gene="F22B5.4"
 /note="cDNA EST yk304c8.3 comes from this gene; cDNA EST yk304c8.5 comes from this gene; cDNA EST yk594h3.3 comes from this gene"
 /codon_start=1
 /protein_id="CAA90356.1"
 /db_xref="GI:3876228"
 /db_xref="SPRMBL:Q19708"
 /translation="MSLSKAPRKTRERFVLSRNASTSNVSPARIQLKPAEAGHFQY
 SRNWSRDPREKVVATQKGTTPQYFLVIRGHAYEYPLVLAAMEVLFCSASVYFG
 KAEIWLDRSNSKAPMDWERLRTYKMPVARDDDGRITKRELMEQDLEAAK
 RCTR"
 complement(10849..12154)
 /gene="F22B5.5"
 complement(join(10849..10984,11033..11343,11397..11521,
 11578..11832,11917..12154))
 /gene="F22B5.5"
 /codon_start=1
 /protein_id="CAA90357.1"
 /db_xref="GI:3876229"
 /db_xref="SPRMBL:Q19709"
 /translation="MNCQOSAKSVKIVGFGFNKASSRCSSEKLIIPKDLTKVNPTS
 EQIKFOKLQVGNNDILRLKETWKNYQLYLCHRKWSTAAICVNFNSVMSIYV
 SDSITVSGQDSITKSLNRIIDPQFGISYQDNTNLLPHLMHMAIDPQVKNYS
 GGVDSKYAAVMKVLTKVRGPTVYVSHKPDVEPELEEDVYKFSQVPEHVYK
 YDSITPPEKKEYILATKFNKVGRAVNAKSGEYIGLGVSTESSGCEIGPIICDNK
 AAQAFIQTIDIPMKDKVILVRCOKPESASWTWIRPFLRRSHEMTPEAHIKENRVI
 PDGLNFSKVFVSSNPSNAPC"
 complement(12240..14032)
 gene

```
CDS
/gene="F22B5.6"
complement(join(12240..12394,12467..12541,12691..12784,
13262..13392,13465..13524,13876..14032))
/gene="F22B5.6"
/codon_start=1
/protein_id="CA90358.1"
/db_xref="GI:3876230"
/db_xref="SPTREMBL:Q19710"
/translation="MKNPYGTFAVNSQCLSTWISQKRRSYFLEALILPTVPSMWL
LAOYHIHSSNIISNMNLNSRGIIVNENATNHFGLCTVTAFLCILSLIVDLI
KTHFETRYVFMIAVACGLFPTVGLIIVIPAKMSITVCSGOLGIOPLTIS
NSTGEERYKLESAGCEFAISSEFNKVFEDLLIQVFIPLHILFLFAFWTSGLRPRM
HEPPE"
15384..21812
/gene="F22B5.7"
/join(15384..15575,15622..16479,16527..17105,18608..18694,
18794..18946,18996..19449,19497..19843,20042..20554,
20672..21600,21677..21812)
/gene="F22B5.7"
/notes="limited similarity with the elongation factor EF-3
(Swiss Prot accession number P25997); cDNA EST EMBL:CI2588
comes from this gene; cDNA EST yk349b7.5 comes from this
gene; cDNA EST yk420b2.5 comes from this gene; cDNA EST
EMBL:M89075 comes from this gene; cDNA EST EMBL:D36040
comes from this gene; cDNA EST yk42id3.3 comes from this
gene; cDNA EST yk42id3.5 comes from this gene; cDNA EST
yk396b3.3 comes from this gene; cDNA EST yk396b3.5 comes
from this gene; cDNA EST yk359g5.5 comes from this gene;
cDNA EST yk359g5.5 comes from this gene; cDNA EST
yk351a5.3 comes from this gene; cDNA EST yk351a5.5 comes
from this gene; cDNA EST yk355a11.3 comes from this gene;
cDNA EST yk355a11.5 comes from this gene; cDNA EST
yk348b2.3 comes from this gene; cDNA EST yk348b2.5 comes
from this gene; cDNA EST yk297a7.3 comes from this gene;
cDNA EST yk297a7.5 comes from this gene; cDNA EST
yk251h3.3 comes from this gene; cDNA EST yk251h3.5 comes
from this gene; cDNA EST yk200h11.3 comes from this gene"
/codon_start=1
/protein_id="CA90359.1"
/db_xref="GI:3876232"
/db_xref="SPTREMBL:O61442"
/translation="MSNWDYLDVDDILPKLPNFDLRESKWKQERKEALLKVL
DNRLSTKASYALGLIHLVAKDANINQOALACIGFAFGAELAKESPSPGLLP
VIEKKMKKPKMLREPLVDCSNVCRPMOSLEFGEDIIALAKPNPOIQOALTYA
ROLDIVPAKOPKPGFIKAVVPFGKLGDDQVREKSLDGLAVORITIDKNKNIL
GDSDBEGKMKKIGYEAKSTSPSAEQANAPVAPVPTSPSAASGSPSGTATA
VSSGAPVAEADPMDLEADVLSKMPDGDITLIESKQERKEALEGLIOLITANPK
LDRKANYGALVERLOKLEKDANIINVAALANCIITGIANCLRPDPFNASVPIIPE
KFEKKPTRLRDPVACIDAVATTNLEAVEIYLAALGKNRPKIKOTDLELDRCPMK
```

Query Match 19.18; Score 339.2; DB 34; Length 25823;

Best Local Similarity 63.38; Pred. No. 5.1e-49;

Matches 633; Conservative 0; Mismatches 223; Indels 134; Gaps 2;

```
Qy 621 acgtaaatcaactgcaaaagctgatatcatcgcctaattcaagggtggtgatcaagt 680
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3064 ATGAACACCTTTTGACAACTTACATTTCTTCCAAACCTTCCAAAGTTGAGCTCAGC 3123
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 681 tcgtacatcaatgatattcttcggtctcgtcagacatcttcttgagtaaacgaattgcgc 740
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3124 ACGGACATCAAGAAATATTCTCCATTACTGGAGCGGATCTCTTCTTCAAACTCGAAGAA 3183
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 741 agctgcgcagcaggttttcgctgcgcgcacacactgttttaacagctcgaatcccttggttc 800
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3184 CTGAGCACTTTGGTTACTTGC-----TTAACTGCTCCAAATCTCTTGCTGCTC 3231
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 801 tgaacatgtgtgacacatctgcatattgtgtcttctaagttaactgatactgacatgca 860
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3232 AGAGCAGGTTGGGGGAGACACACCTCTATTGTGTTCTTAAACGTAAATGGAATCTGCA 3291
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 861 tgcataagaaaagctgtgatccgcataattgatacgtgagctccttcgcgaagcat 920
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3292 TTGATGAGAAAAGTTGAGAACGATCAGCGTATTGTGTAGACGTGAGACCTCTGTTCAGCAT 3351
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Qy 921 taactctgtgatataatccaaattatagcaaatattatcaagaagcacatccatcagc 980
||||| ||||||| ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3352 CAATCACTTGGATATTTCCAAATTTGTTAGCAGAACACTTATCAAGACACACCGTTCTC 3411
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 991 attgaatcttccacagatcaacgcgttaccatcatcgcgaacgaagcaatggaacccgc 1040
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3412 ATCCAAATTTGACAGCGATATCACCATTTCCATCATCAACAGTGCAGAGTGAACGACAGC 3471
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1041 gcaagaataatcaacggtttgaagatctgcagatgcatattatgataaactgctgaccaa 1100
||||| ||||||| ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3472 GCAGAAATATCAACGCTTTCAGATTCACAGCTCCATTTGTGATAGACTTGTGTCCAAT 3531
||||| ||||||| ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1101 gatacgaatggaactgtgtaccggttgcaccatccatccaaattcatcaagcagatc 1160
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3532 AGTTGCGAATATGAACTGCTTCTCCAGTTGCTCCCTTAAGATCTCGTATTATACAGAT 3591
||||| ||||||| ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1161 tgcgacgtggacaatttgagtttgaaaagcagttgtgatttcagatacctcaatctgtgc 1220
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3592 TGGCATTGGAACACCTGGCTTTGGAAAGCGGTGTAAGATCAAGACTTCAATTTGGGT 3651
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1221 actaactgtttatcagcttccatgaaagcattgatactgatatgca----- 1270
||||| ||||||| ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3652 AGAGACTGTCTTGTGGCTTCCATGTAGAACGATTTGAACACGATGACGATGGAATTTCA 3711
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1270 ----- 1270
Db 3712 ATTTCATTTTAAACGAGAAAGCTTCTAGTTGATATCAACCTTTATATTTCCGTATCC 3771
||||| ||||||| ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1270 -----cgatcaact 1278
Db 3772 TAAACCTTCAAGAACATTCCTTAAATAATACAGAACTTTCATGCTTACCGATTAAC 3831
||||| ||||||| ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1279 ttggttaacaaataatgataagaaacgaatgcaacagttgtgttacaataatccacagt 1338
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3832 TTGGTGCACAAATTTGGCGATGGAAGACGACGACAGCTGTGGTAACGAACATCCCTTT 3891
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1339 ggaattcagagatcgttgatcgacgcacacatgatgatcaaaatggaagtgaatccgcga 1398
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3892 GGATTCAGAGATCTGTGATGAGCAACGTTCAAGTATCGAATGATGATGATGATTCAGCG 3951
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1399 actgaagctcacccttcatcatcttcgcgaacactcttgatcataaagaccttcatata 1458
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3952 ACTTGACGACCTCTCTGCTATTTCTCATTTCTGTTGATGGAAGAGTCCCTTAACGTAG 4011
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1459 acatgcttcggaatgcatcactgctgtttaaacttgattg 1498
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4012 ACGTGCTCCTCGAAGCATTAAGTGTTGTTGAGATTACTG 4051
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 2
CEC47G2
LOCUS 34831 bp DNA INV 02-SEP-1999
DEFINITION Caenorhabditis elegans cosmid C47G2, complete sequence.
ACCESSION Z49125.1 GI:790365
VERSION 249125.1 GI:790365
KEYWORDS HMG; Cuticulin; Fork head.
SOURCE Caenorhabditis elegans.
ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditiida;
Rhabditiina; Rhabditiolidea; Rhabditiidae; Pelodermatidae; Caenorhabditis.
REFERENCE
1 (bases 1 to 34831)
Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
Bonfield,J., Burton,M., Connell,M., Copsey,T., Cooper,J.,
Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
Fullon,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
Saunders,D., Showkeen,R., Smaildon,N., Smith,A., Sonhammer,E.,
Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
Wilkinson-Spratt,J. and Wohldman,P.

Query Match	Best Local Similarity	Matches	Score	DB	Length
14.0%;	69.7%;	352;	249;	34;	5792;
Conservative	0;	Mismatches	150;	Indels	3;
				Gaps	1;
651	atcgcgaattcaagaaggtgtgataatcagttcgtatcatcatatgatatctccggtttctgc	710			
4080	atcactaatgtctgacagggcctgaatccggtgctgacatcaaaaggctgtgtactagtcacc	4021			
711	agatctttcttgtagtaaggaagtgtgcagctgtgcagcaggttttctgtgc---	ggcacc	767		
4020	cccactcttttggatcagcgggaggttgacgcccgtctgttttttgcgcgtatggaacgc	3961			
768	acctttttaacagctccgaatccctgtgtgtctgaaacatgtgtgtgcgaacacattgcct	827			
3960	acctgagagcagacagcccaaatcctttggggcctacatgcgaatggsgtctggggcattccgt	3901			
828	attgtgtctttaaattgtgtaaacatgtatctgtgacattgtatagaagaagctgtatgcattccgc	887			
3900	gtttggcctcttttaaggatagatgtatctgtgacatgcataaagaacgtgtgactgactccgc	3841			
888	atatgtgatacagtagcttctgtccgagccatataatcgttggatatgttccaaattat	947			
3840	atattttgtagacgttccgcttcttggccggccatcaaatcgggggactgacaggtgtt	3781			
948	tagcaaatatlatcaagaagacatccatccatcagcatttagaatttccacagatcacgct	1007			
3780	cagcaggtattttgttaacggcagacacacccctccgttgtagagatctgacatccatccgct	3721			
1008	accatcatcgacaagaagcagaatgtgacaacccgcagagaagatcaacggtttcagaatc	1067			
3720	gccatcatcaacaaagcagatgtagacagacagacagaaacgattccacagattccgaatc	3661			
1068	gcattgcatcattatgataaactggtgtgcgaatgatgatgcaaatgtgaactgtgtgacggt	1127			
3660	gcaaatctcactttgtgtacttgcgtgacgcatgattgacgaactgattacggtgtcccggt	3601			
1128	tggtccaccatccaaaattcataa	1152			
3600	tggacctccgtcgagaatctgtaaa	3576			
RESULT	6				
LOCUS	MAATCUT1/c				
DEFINITION	M.arteillella Mcut-1 gene.				
ACCESSION	X96677				
VERSION	X96677.1				
KEYWORDS	cuticle protein; cuticlin 1; Mcut-1 gene.				
SOURCE	Meloidogyne arteillella				
ORGANISM	Eukaryota; Metazoa; Nematoda; Secernentea; Diplogasteria; Tylenchida; Tylenchina; Tylenchoidea; Heteroderidae; Meloidogyinae; Meloidogyne				

[illegible]


```

gene
/clone="F53f1"
/complement(14923..16486)
/gene="F53f1.1"
complement(join(14923..15347,15663..15774,15835..15984,
16033..16271,16318..16486))
/gene="F53f1.1"
/note="predicted using GeneFinder; similar to cuticlin"
/codon_start=1
/codon_id="CAB03124.1"
/db_xref="GI:3877456"
/translation="MDNKEIRRLDEYFPPKFLISESSKYDCHLEDMHLVQLLLATL
ANVPDPNLSLYGVVECDSTRTSVQIKTEKPVGVIFVDFRASEEVCISRGSRIS
AFLEIEIGCALRORLNPKNGLARTTITSEHPYFITYKDRYNNLCYRESQVIV
ANNISVDEISTISINVLMTWCTCYOILISGGEPEVEFLIGQOYVHOKMCDKGM
IILIKEDFCVHVHICVDGDRGTSFLIDNSGCSIDKFLISNLEYGNLGOEAAHY
YKPADRALPEFCOISITVEKPEDECVARPIEDEVEGGCAPVPGPPYGMISKNLVNIH
QKNSFLTFKSPMYVVGITNYHNKAIOKL"
17396..18524
/gene="F53f1.2"
join(17396..17599,17683..17926,17977..18146,18197..18286,
18339..18524)
/gene="F53f1.2"
/note="similar to aldehyde reductase; cDNA EST yk473d8.3
comes from this gene; cDNA EST yk473d8.5 comes from this
gene"
/codon_start=1
/codon_id="CAB03127.1"
/db_xref="GI:3877459"
/translation="MSDSFANVPGSGSKLNSGYNIPVFGCTYKVTGENVPDAIAL
TAGYRMPPTAKYILNEKEGELKILPPLHSGRSDDVFLTSKFPPESKCRACGAFV
ESLDSLOTDTYIDMLVHYPRKNDSDNDVNNAEYKIAEYLEAKAGKRSIGVS
NEIYHLELKYTAHVPCANDLEHHPHARIPLOKCKEKNIFQAFSESLARHEPKL
IEDPVVLELAKHNTSVPLVLLAMLRQNVGIVPSVPSRIVENFKVIDIALTPEDI
GSLFALDRGOHYIRCTGMLVK"
18714..20251
/gene="F53f1.3"
join(18714..18887,18935..19056,19113..19302,19353..19448,
19919..20008,20072..20251)
/gene="F53f1.3"
/note="similar to aldehyde reductase"
/codon_start=1
/codon_id="CAB03128.1"
/db_xref="GI:3877460"
/translation="MSIKLNTGIDCPILGCTYKITGDVLPYLDALTAGYRLPDA
KYVNNKEIGDALEILPKHNLKREDIFITFKHPNTVEVKKLVDEBSLSLTKSYID
MYLIHPKSPDYGDODPMNKTIRIATWMDMECKNAGKIRSVGVSSFEIHLDELDL
GRKPPCCNVEYHPHTRELKNYCKSEGIFFQAFSESLARHETLISSEITIRLAEK
YHPKTTYLLSNATISQKVGIIIPKSTNERLQNLKTVLLEEEVKKIKCINLIDKHVYR
TTGWLVLV"
21582..22009
/gene="F53f1.4"
join(21582..21670,21781..22009)
/gene="F53f1.4"
/note="predicted using GeneFinder; similar to cuticlin;
cDNA EST EMBL:D72018 comes from this gene; cDNA EST
EMBL:D74799 comes from this gene; cDNA EST EMBL:C11739
comes from this gene; cDNA EST EMBL:C12451 comes from this
gene; cDNA EST EMBL:C10189 comes from this gene; cDNA EST
EMBL:C10682 comes from this gene; cDNA EST yk506b3.3 comes
from this gene; cDNA EST yk437e9.3 comes from this gene;
cDNA EST yk437e9.5 comes from this gene; cDNA EST
yk311h10.3 comes from this gene; cDNA EST yk311h10.5 comes
from this gene; cDNA EST yk341f6.3 comes from this gene;
cDNA EST yk300g2.3 comes from this gene; cDNA EST yk300g2.5 comes
from this gene; cDNA EST yk234b3.3 comes from this gene;
cDNA EST yk234b3.5 comes from this gene; cDNA EST
yk191d9.3 comes from this gene; cDNA EST yk191d9.5 comes
from this gene; cDNA EST yk677d4.3 comes from this gene"
/protein_id="CAB03125.1"

```

gene
CDS

Query Match

```

/db_xref="GI:3877457"
/translation="MOKIYFFPAIAVSOAFLPSPGGGCGCAPPPPPPCGCGPS
LPEQLPREFELPRLSLPFGCGCGGCPACGGAAPYVAPPAGVAAAPPAGGYATA
GRK"
25572..26408
/gene="F53f1.5"
join(25572..25660,25850..26408)
/gene="F53f1.5"
/note="predicted using GeneFinder; similar to cuticlin;
cDNA EST EMBL:T01107 comes from this gene; cDNA EST
CEESN60F comes from this gene; cDNA EST EMBL:D70966 comes
from this gene; cDNA EST EMBL:D71580 comes from this gene;
cDNA EST EMBL:D72120 comes from this gene; cDNA EST
EMBL:D72428 comes from this gene; cDNA EST EMBL:D73526
comes from this gene; cDNA EST EMBL:D74402 comes from this
gene; cDNA EST EMBL:D74922 comes from this gene; cDNA EST
EMBL:D69807 comes from this gene; cDNA EST yk505b11.3
comes from this gene; cDNA EST EMBL:D76322 comes from this
gene; cDNA EST EMBL:C12136 comes from this gene; cDNA EST
EMBL:D69807 comes from this gene; cDNA EST yk480h10.5 comes
from this gene; cDNA EST yk492b11.5 comes from this gene;
cDNA EST yk480h10.3 comes from this gene; cDNA EST
yk480h10.5 comes from this gene; cDNA EST yk445a7.3 comes
from this gene; cDNA EST yk445a7.5 comes from this gene;
cDNA EST yk440b8.3 comes from this gene; cDNA EST yk440b8.5
comes from this gene; cDNA EST yk377b11.3 comes from this
gene; cDNA EST yk377b11.5 comes from this gene; cDNA EST
yk473g12.3 comes from this gene; cDNA EST yk473g12.5 comes
from this gene; cDNA EST yk459g4.3 comes from this gene;
cDNA EST yk459g4.5 comes from this gene; cDNA EST
yk482d10.3 comes from this gene; cDNA EST yk482d10.5 comes
from this gene; cDNA EST yk455c6.3 comes from this gene;
cDNA EST yk455c6.5 comes from this gene; cDNA EST
yk417a12.3 comes from this gene; cDNA EST yk417a12.5 comes
from this gene; cDNA EST yk414h1.3 comes from this gene;
cDNA EST yk401f4.3 comes from this gene; cDNA EST yk401e5.3
comes from this gene; cDNA EST yk401e5.5 comes from this
gene; cDNA EST yk399e6.3 comes from this gene; cDNA EST
yk399e6.5 comes from this gene; cDNA EST yk392h1.3 comes
from this gene; cDNA EST yk392h1.5 comes from this gene;
cDNA EST yk391b10.3 comes from this gene; cDNA EST
yk391b10.5 comes from this gene; cDNA EST yk342c6.5 comes
from this gene; cDNA EST yk342c6.5 comes from this gene;
cDNA EST yk345c4.3 comes from this gene; cDNA EST
yk345c4.5 comes from this gene; cDNA EST yk307f5.3 comes
from this gene; cDNA EST yk307f5.5 comes from this gene;
cDNA EST yk258a5.3 comes from this gene; cDNA EST
EMBL:C09046 comes from this gene; cDNA EST yk305d6.5 comes
from this gene; cDNA EST yk314h1.5 comes from this gene;
cDNA EST yk331e10.5 comes from this gene; cDNA EST
yk343g4.5 comes from this gene; cDNA EST yk375f1.5 comes
from this gene; cDNA EST yk404c1.5 comes from this gene;
cDNA EST yk447c9.5 comes from this gene; cDNA EST
yk447e5.5 comes from this gene; cDNA EST yk488e6.5 comes
from this gene; cDNA EST yk382a5.3 comes from this gene;
cDNA EST yk506h10.3 comes from this gene; cDNA EST
yk507f12.3 comes from this gene; cDNA EST yk518b6.3 comes
from this gene; cDNA EST yk530f12.3 comes from this gene;
cDNA EST yk536e1.3 comes from this gene; cDNA EST
yk537c8.3 comes from this gene; cDNA EST yk540b5.3 comes
from this gene; cDNA EST yk545b10.3 comes from this gene;
cDNA EST yk557b11.3 comes from this gene; cDNA EST
yk562h11.3 comes from this gene; cDNA EST yk566f10.3 comes
from this gene; cDNA EST yk596a10.3 comes from this gene;
cDNA EST yk596e11.3 comes from this gene; cDNA EST yk614f1.3
comes from this gene; cDNA EST yk618d8.3 comes from this
gene; cDNA EST yk620c8.3 comes from this gene; cDNA EST
yk620d5.3 comes from this gene; cDNA EST yk622e4.3 comes

```

6.0%; Score 107.4; DB 34; Length 39478;

Best Local Similarity 61.7%; Pred. No. 1.5e-09;
Matches 171; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 781 gctccgaatcttggtggtcgaacatggtgctgaacacatgctcttggctctta 840
Db 15074 GCTCCACCGCCTTCTACATCTTGCATATGTGCTCCACGCAITCCTGCTGCTTTTG 15133
QY 841 atgtaatacgcattgcatgcatgataagaagctgcatccgcacattggtatagc 900
Db 15134 ACGGAATATGATATCTGACTGGAAGAAAAGAGATCTCGATCGACAACTTATACCA 15193
QY 901 tgaactcttgccagccattaaatcgttgatataatccaaattattagaatatta 960
Db 15194 TGAGCTTCTTCCACGACGAGTAAGTTTCCCGATCTCAAGCTTCGACATAGAACTTA 15253
QY 961 tcaagagacatccatgcatgcatgataatccacagatataccgtccatcatagaca 1020
Db 15254 TCAATTTGACACCCCTTCTGAGCTATGGAANAATGAGGTTTCACTCTCATCGTCAACA 15313
QY 1021 aagcagaatgacacacccgcagaaatcatcaacg 1057
Db 15314 GAACACGTGTGACTACATGCAGATGAGTCTCTG 15350

RESULT 10
CE04D5 31536 bp DNA INV 02-SEP-1999
LOCUS Caenorhabditis elegans cosmid E04D5, complete sequence.
DEFINITION Z66496
ACCESSION Z66496.1 GI:1041307
VERSION HMG; Cuticulin-like protein.
KEYWORDS Caenorhabditis elegans.
SOURCE Eukaryote; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
ORGANISM Rhabditia; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 31536)
AUTHORS Wilson, R., Alnseough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkes, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Latrelle, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rikfen, L., Roopra, A., Saunders, D., Showkeen, R., Smaldon, K., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sporat, J. and Wohldman, P.

TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans

JOURNAL Nature 368 (6466), 32-38 (1994)

MEDLINE 94150718

REFERENCE 2 (bases 1 to 31536)

AUTHORS McMurray, A.

TITLE Direct Submission

JOURNAL Submitted (25-OCT-1995) Louis, MO 63110, USA. E-mail: jss@sanger.ac.uk or tw@nematoe.wustl.edu

COMMENT Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.

For a graphical representation of this sequence and its analysis see:

http://webace.sanger.ac.uk/cgi-

bin/displayrdb-wormaceclass-Sequence&object=E04D5

Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

IMPORTANT: This sequence is not the entire insert of clone E04D5.

FEATURES

source
gene
CDS

It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
The true left end of clone E04D5 is at 1 in this sequence. The true right end of clone E04D5 is at 2110 in sequence Z48585.
The true left end of clone ZK673 is at 31437 in this sequence. The true right end of clone T09F3 is at 11412 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence Z49070.
The end of this sequence (31437..31536) overlaps with the start of sequence Z48585.

Location/Qualifiers

1..31536

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="II"

/clone="E04D5"

6082..8326

/gene="E04D5.1"

join(6082..6250,6307..6379,6432..6509,6564..6696,6750..6930,7003..7283,7337..8179,8228..8326)

/note="CDNA EST EMBL:D71941 comes from this gene; CDNA EST EMBL:D74691 comes from this gene; CDNA EST EMBL:D65192 comes from this gene; CDNA EST EMBL:D68540 comes from this gene; CDNA EST EMBL:C11194 comes from this gene; CDNA EST EMBL:C13213 comes from this gene; CDNA EST EMBL:C11881 comes from this gene; CDNA EST YK412d7.3 comes from this gene; CDNA EST YK412d7.5 comes from this gene; CDNA EST YK459b6.5 comes from this gene; CDNA EST YK454d9.3 comes from this gene; CDNA EST YK454d9.5 comes from this gene; CDNA EST YK350b11.3 comes from this gene; CDNA EST YK350b11.5 comes from this gene; CDNA EST YK413f5.3 comes from this gene; CDNA EST YK413f5.5 comes from this gene; CDNA EST YK402d12.3 comes from this gene; CDNA EST YK402d12.5 comes from this gene; CDNA EST YK375c1.3 comes from this gene; CDNA EST YK375c1.5 comes from this gene; CDNA EST YK374f8.3 comes from this gene; CDNA EST YK374f8.5 comes from this gene; CDNA EST YK365e8.3 comes from this gene; CDNA EST YK365e8.5 comes from this gene; CDNA EST YK256f12.3 comes from this gene; CDNA EST YK256f12.5 comes from this gene; CDNA EST EMBL:C12803 comes from this gene; CDNA EST EMBL:C12996 comes from this gene; CDNA EST EMBL:C13136 comes from this gene; CDNA EST YK268c5.3 comes from this gene; CDNA EST YK519a6.3 comes from this gene; CDNA EST YK533e1.3 comes from this gene; CDNA EST YK556e6.3 comes from this gene; CDNA EST YK578a6.3 comes from this gene; CDNA EST YK598a11.3 comes from this gene; CDNA EST YK640h2.3 comes from this gene; CDNA EST YK650h4.3 comes from this gene"

/codon_start=1
/protein_id="CA91279.1"
/db_xref="GI:3875451"
/db_xref="SRRMBL:019052"
/translation="MRPNTTPSRDSSEFYISSFSFLLLPKFTTSPRLSOTLI QINKGNDLVAVRSSSEGLKRLGLGDAVTEQNTSDVACNFAVSNQDLAY CNDQVREVEFLIATNKELICVELKRTKILFSPDNFLIPEPAVVGPKAEQKEP NRVVYSLADSKHVTESAPKEASMEPOFSDEDLARMSVEFEFTNMEEDYDKL VEGATNFLESGPAPRHVAVVYPAVSGAPRAVRHVHVSFPVVCNRPFEKDKAM TNNOROSLIILASVEUKDNOSYVEBOSYLINTLNGESVYVPLEKGTIYAKNP NGRFVATCYTMEPAKVTFTNPRGVPIEDTIEGRNDVFNVAENIVLICFGNIARK MERFVETKEKEIISIEVPNTTLDPMADGQFVYCTTAPLRIDNSYREMHYARLA ETHFESPEKLEIEMVRMPNCTGYNKFAIKELTKTKMAAGLPIRKDSHPINNVPAVAY RQAGAYIPLHLRKLPGGSGAGPSAAAPPGQONORPADPRANGCNAPQPRPOOS EORRKAFLKKKVEIKVKORVANDQLOPQNMERTIQRENVLELSLKTII"

complement(8619..11718)
/gene="E04D5.2"

complement(join(8619..8761,8813..8988,8995..10028,10089..10283,10330..10614,10730..10935,11582..11629,11691..11718))

```

/gene="E04D5.2"
/codon_start=1
/protein_id="CA91282.1"
/db_xref="GI:3875454"
/translation="MGENLFYADIHSTIPEKCKLKI RVELSEMAOKHLFGLKLP
INPOLAMFRRSLPSSIFLPLATIDMLICTYTLFGVDAGIWRKNTLFLYHR
YIVFVETAKVVOFAIPILITIPFIRKLVAKRENNPILIRGAGTSTEVKASPVG
YDFHVTAIOMMFPPVLLNLNTITIKLVAEKRENNPILIRGAGTSTEVKASPVG
NLPNVYLQVAADVIKESLIHRSRSRSQLRNAITMIAIVSYLCKNVHLPLTI
LEREDPSVLYESTRMOSEFYIVLSDVTSICYMASAIRIFIAKCNPKLRQETITDI
IKRKSTFTNNS"
25834..29213
/gene="E04D5.3"
/join(25834..25873,27277..27512,27582..27721,27768..28110,
28158..28410,28491..28811,29092..29213)
/gene="E04D5.3"
/feature="similar to cuticlin like protein"
/codon_start=1
/protein_id="CA91280.1"
/db_xref="GI:3875452"
/translation="MPSRLAMKELKGGEPEVCEFTASISLFTKRNSENGKAVKGY
VSEPCMTVGCGKTHREVRHDSGVRROREINGVISATVILISFHSIFIKIDVAY
RVSCFTVESTKRVHNVHDSALTQLESETPVCRREILNEAGSGPIKRYRIDVY
YHKTVCVALENNVCMKHSCTVYSGOGPPYTVIDANGCSGVIIOLNLYTDLTA
GTLAPVFKFADAGLYFNCQIQLTIKDVNTGCSNTPQCPSPQYVVERPKETPAEP
YVDHESGYPTRPANYPVASRRYPPTPTQADASVPSAPPAGADIDNGPEPQPI
YIAETPENAYDIVEGNDTEOPFTTSAAYTEGVSRLIKRVVESTEDVSNKRP
VIVGDIIDPERGILVEGMEDEGTETNGATPALREARNNSOEKTCFSTSRMYFT
LILCLLFTVTVVPIVIVQKROLIAQAFKRP"
29675..31348
/gene="E04D5.4"
/join(29675..29840,31127..31214,31267..31348)
/gene="E04D5.4"
/codon_start=1
/protein_id="CA91281.1"
/db_xref="GI:3875453"
/translation="MWSINLVHILLVTFSSVHVTVAVTKATGETTVRAGAGODIGD
VSSSFYEETTTASTCADDPNTDCTQYTFLOSNKRYPLLOQFCATKCGCGSGSTIAP
VACVDNSTQ"
BASE COUNT 10458 a 5592 c 5636 g 9850 t
ORIGIN
Query Match 4.7%; Score 83.8; DB 34; Length 31536;
Best Local Similarity 56.1%; Pred. No. 1.6e-05;
Matches 179; Conservative 0; Mismatches 137; Indels 3; Gaps 1;
QY 819 acattcgctatttgctttaaactgaatctgcatcgcatgataaagctgtga 878
28101 ACATCCATAATTTAGCTTGTGATTGTAATTGACAAATTGACAAATGAAGCCACGC 28042
QY 879 tgcattcgcatattgttatactgtagctctcttgccagcatiaaactgtgtgataltc 938
28041 TTTGTCACGCAATTTGAAGACTGGTGGCAGATTTCACACTGTCAAGTCAGAGGTCTACTC 27982
QY 939 caaatatttgcaaatatttatacaagagcacatcat---cagcaattagaattccac 995
27981 AAGATTTCGAAGATCACCCTCATACGAGACATCATTCATGCAATGACAGTTACTGG 27922
Db 27981 AAGATTTCGAAGATCACCCTCATACGAGACATCATTCATGCAATGACAGTTACTGG 27922
QY 996 agtataccgcttaccatcatcgacaagaaggaatgagacaacccgcgcagaagtatcaac 1055
27921 TGGTCCACCTTGTCCATTAACCGTACATGATGATTCATACATGATACATATTCCTC 27862
Db 27921 TGGTCCACCTTGTCCATTAACCGTACATGATGATTCATACATGATACATATTCCTC 27862
QY 1056 ggttcaagaatcgcatgtcatttataaactgctgacgaatgataatgcaaatgaaac 1115
27861 AAGTTCAGCAAGCAAGTTCATTGTGATGACTGATCTCCGATTCCAGCGCTACTTAAT 27802
QY 1116 tggctgacgcggtgtgtcca 1134

```

Db 27801 AGGTGATCCACCGGCTTCA 27783

RESULT 11
 CEF53B6
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS

TITLE
 JOURNAL
 MEDLINE
 SUBMITTED
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

2 (bases 1 to 32412)
 Nature 368 (6466), 32-38 (1994)
 94150718
 White, S.
 Direct Submission
 Submitted (21-Oct-1996) Louis, MO 63110, USA. E-mail:
 jes@sanger.ac.uk or rtw@nemato.de.wustl.edu
 Coding sequences below are predicted from computer analysis, using
 predictions from GeneFinder (P. Green, U. Washington), and other
 available information.
 For a graphical representation of this sequence and its analysis
 see:
 http://webc.sanger.ac.uk/cgi-bin/display?db=wormacc&class=Sequence&object=F53B6

bin/display?db=wormacc&class=Sequence&object=F53B6
 Current sequence finishing criteria for the C. elegans genome
 sequencing consortium are that all bases are either sequenced
 unambiguously on both strands, or on a single strand with both a
 dye primer and dye terminator reaction, from distinct subclones.
 Exceptions are indicated by an explicit note.
 IMPORTANT: This sequence is NOT necessarily the entire insert of
 the specified clone. It may be shorter because we only sequence
 overlapping sections once, or longer because we arrange for a small
 overlap between neighbouring submissions.
 IMPORTANT: This sequence is not the entire insert of clone F53B6.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we arrange for a small overlap between
 neighbouring submissions.
 The true left end of clone F53B6 is at 1 in this sequence. The true
 right end of clone F53B6 is at 5116 in
 sequence 281523.
 The true left end of clone F32H2 is at 32306 in this sequence. The
 true right end of clone F08G11 is at 7884 in this sequence. The
 start of this sequence (1..105) overlaps with the end of sequence
 280220.
 The end of this sequence (32306..32412) overlaps with the start of
 sequence 281523.

FEATURES
 SOURCE

1..32412
 location/Qualifiers
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /chromosome="I"
 /clone="F53B6"
 1480..3649
 /gene="F53B6.1"

	Db	29738	GGAATGAAACAGAGGTGCATTCATTGTGAAGCAAAAATGCATAATTTCCACGGTGCATGCCAGTGCTCST	29797
	Oy	1354	gtacgcgaaccatcattcatgataagaatggaaagtg	1387
	Db	29798	TGGCTCGCATATTTGCATGCTCCAATTGGAATWG	29831
RESULT	12			
CeYl11B2_0/c				
MPCOMMENT				
Sequence split into 7 fragments			LOCUS CeYl11B2 Accession Z98657	
Fragment Name	Begin	End		
CeYl11B2_0	1	110000		
CeYl11B2_1	100001	210000		
CeYl11B2_2	200001	310000		
CeYl11B2_3	300001	410000		
CeYl11B2_4	400001	510000		
CeYl11B2_5	500001	610000		
CeYl11B2_6	600001	614429		
LOCUS	CeYl11B2	614429 bp	DNA	HTG
DEFINITION	Caenorhabditis elegans chromosome III clone YL11B2,	** SEQUENCING		
IN PROGRESS ***.	In unordered pieces.			
ACCESSION	Z98857			
VERSION	Z98857.36 GI:6065761			
KEYWORDS	HTG; HTGS; PHASEI.			
SOURCE	Caenorhabditis elegans.			
ORGANISM	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilitia; Rhabdilidae;			
	Rhabdilitina; Rhadilicoidea; Rhabdiltidae; Pelodermaeae; Caenorrhabsitis.			
REFERENCE	Submitted (27-OCT-1999) Nematode Sequencing Project, Sanger Centre,			
AUTHORS	Hinxton, Cambridge CB10 1HQ, UK and Department of Genetics,			
JOURNAL	Washington University, St. Louis, MO 63110, USA. E-mail:			
	jess@sanger.ac.uk or rwenemate@wustl.edu			
	On Oct 19, 1999 this sequence version replaced g1:6016917.			
COMMENT	Order of segments is not known; 800 n's separate segments.			
	IMPORTANT: This sequence is unfinished and does not necessarily			
	represent the correct sequence. Work on the sequence is in progress			
	and the release of this data is based on the understanding that the			
	sequence may change as work continues. The sequence may be			
	contaminated with foreign sequence from E.coli, yeast, vector,			
	phage etc.			
	* NOTE: This is a 'working draft' sequence.			
	* This record will be updated with the finished sequence			
	* as soon as it is available and the accession number will			
	* be preserved.			
FEATURES	LocationQualifiers			
source	1..614429 "Caenorhabditis elegans"			
	/organism="Caenorhabditis elegans"			
	/db_xref="taxon:6239"			
	/chromosome="III"			
	/clone="YL11B2"			
BASIS COUNT	193560 A 103068 C 106254 G 187431 T 24116 others			
ORIGIN				
	Query Match	4.0%; Score 71.6; DB 32; Length 110000;		
	Best Local Similarity	51.8%; Pired. No. 0.0017;		
	Matches 164; Conservative	0; Mismatches 154; Indels	0; Gaps	0
Oy	1 ttttcttttttttttttttttttcacaatatgatcaaatccatttaagtccaataatgccaaa	60		
Db	73408 ttttttttttttgttctaTTTTTTTGTACTAATAAATGATTTTTTAATAAAAAATGAAAAAAAAA	73349		
Oy	61 taagtgaattigttagcgactgaatgatgbcccataaanaatlgyggcaalacaaaggat	120		
Db	73348 TGAAAAAAAAATGTAACAACAAAAAACGAAAAAATGCAAGAAGAAAAATGAGAAAAATGAGAAA	73289		
Oy	121 aaatgtaataaacacacaaatgtaaacyctgttagaataaaaaatgacgaatggttatatg	180		

Db	73288	AAATTGAGAAGAAATGAAAAAAAAATGAGCAAAAAAATGAAAAAAAAAAAAAAAAATGCAAAAAAAATG	732229
QY	181	atgataatacaacgaaactlcatgacctaatgaaacattcaatgtctgca	240
Db	73228	AGAAAAAAATGAAAAAAATGACAAAAAAATGAGAAAAAAATGTAAGAAAAAAATGTA	73168
QY	241	tataaagactactgataatgaatcaagcttccgaagcgtataaagcgatagaagaag	300
Db	73168	AAAAAAAAAAGAAAAAATGACAAAAAATGACAAAAAATTGAAAAAAGGCGCACATC	73109
QY	301	tttggatgaatatccac	318
Db	73108	TTTTCGTGGCAAAACCAC	73091

[illegible]

```

Available information:
  for a graphical representation of this sequence and its analysis
  see:
  http://webace.sanger.ac.uk/cgi-
  bin/display?db=wormace&class=Sequence &object=Y18D10A
  Current sequence finishing criteria for the C. elegans genome
  sequencing consortium are that all bases are either sequenced
  unambiguously on both strands, or on a single strand with both a
  dye primer and dye terminator reaction, from distinct subclones.
  Exceptions are indicated by an explicit note.
  IMPORTANT: This sequence is NOT necessarily the entire insert of
  the specified clone. It may be shorter because we only sequence
  overlapping sections once, or longer because we arrange for a small
  overlap between neighbouring submissions.
  IMPORTANT: This sequence is not the entire insert of clone Y18D10A.
  It may be shorter because we only sequence overlapping sections
  once, or longer because we arrange for a small overlap between
  neighbouring submissions.
  The true left end of clone P08A8 is at 152766 in this sequence. The
  true right end of clone B0019 is at 105 in this sequence. The start
  of this sequence (1..105) overlaps with the end of sequence
  AL008866.
  The end of this sequence (152766..152878) overlaps with the start
  of sequence Z99710.

```

FEATURES		Location/Qualifiers
source		1..152878 /organism="Caenorhabditis elegans" /db_xref="taxon:6239" /clone="Y18D10A"
gene		complement(7920..26256) /gene="Y18D10A.1"
CDS		complement(join(7920..8081,9752..9924,10879..11164,12369..12699,13906..14204,15129..16655,17748..18104,19860..19949,20008..20088,20709..20807,22205..23127,24486..24744,25724..25909,26020..26130,26236..26256)) /gene="Y18D10A.1" /note="predicted using GeneFinder" /codon_start=1 /protein_id="CAA22308.1" /db_xref="GI:3979929"
gene		/translation="MDEDIAEACASGSCANVKNKNGLLLMONKQRAEKNQDSKVQAECEYLRRKLVTDAIYRESYQREAEINAEINAEASLKQGYSEFTKKSASLQEOADDAEKHKDQANWRNRYDSVCALELEANNIDIKSYESTVKLETAGKRMFLDQVTLTKAKIENSSKTKLIDYKLSVIOVVALSTSNLSKOSKIKDLVVFRRDVRLEFNQPMKNLSLSTSPSISATPEASEDEHDDVITEPAPPIGEOTLSDDDEEEVPEDEADEVDREIROLENLQVRKEKTVRRDPDTAKMETLAKSAPRLVQAEFPVKKMEGIJHKPTEKXSGFLNSTKIHRONPSTDEFTTAQATAEKKALQAGAGASEPGSSSIHGKKGKTVRELQTEMKGEVKKAAKMRKEASGRSQAPAPAPASLQDPQDGLMSDPSDSETEHQKQKHLPAVTRKSARLSALPVTPKXASSSKMPSPSPSTGRRGRRRRTLSTMSMEPAAAVPAPGRPRSAAKAVSENTPLSEA PSAVKRGRRPRSRTMSITDESEPTSTAKRSKASDEEBODLKTNKSPKKPKSKTEETVGDVLKKRLRDATATATVYHNGPRLPRKMERMAPATVYSKKEKPKNAGSADSSINEBEDETMLEEOTLDLPOOTSQEPRIKSGSLDEQDASEHSTVSPAPELTNAPAVEASEASAEPKIDIPEQATPILALALALPVSTALEPKAQNPTALPPTSEISGRAPQLPPTSGSAAPVDLSEILSGAKTKTKTKTKRKAPPAVQKSISSITQOAPPTSVQAPPTSCSAAPVDLSEILSGAKTKTKTKTKMPYDOKKISSEAPTSIDSAPTSVHOOTPKSPQIINSKGLIDSDSEDEEEBEEGMEVEEEERAPRISDSIOASEPSTANVPEKVIYAVKTFSPIDSTSVAPASAPPKAPARKINQALDOOASEPPEPKAKLIDIAKIPKIKTISIPASSSTTSDDLSDIILAGAKTKTKPKAPPTVQOTTRKINAQKRAKSPPTAGTTAPKQYIKGIDS VLPSSSSSTEPSAPDSASTSMKGGGAIMEAIVPCRPGRKAIKREQPIGEMDMONVEKGCGKVKIKTHLQALDLKIPPEELKRMEEKIKLGDSPILTPSPADVEMFEFLRETSAADMAVILNRORIDANLKPKMKEENFLOVSVTLNEHQDLLEFISIIIFELCARSEIGSNMCGRLIRLFGHAVFSESANPDVTSPIVOKKSTWIRRLFOILLNPNQIGKSTCYILMSDAVKCRFLVEEMODHNGSEHILFRVLMHDSQASAVNMLNAKEDVAIVAPTSNDVTACINAHKQCIDDDVTPSLKSTIFLAKTAPAEVLNVAEFLROMKDLIEQFLKPKMEKPKRTIASVTATVTKMTKSAEMLKKEKSKQKNVLRVINDPFESSHAIIPALKOISPOVSQRLIEIIESSIEINQVEMTGEDAIRAVQNLVDLSITYNPLKPTTLPPSS"
gene		complement(26502..29006) /gene="Y18D10A.2"
CDS		complement(join(26502..26573,27346..27405,27490..27580,27624..27788,28732..29006)) /gene="Y18D10A.2" /note="predicted using GeneFinder" /codon_start=1 /protein_id="CAA22309.1" /db_xref="GI:3979930"
gene		/translation="MOMNPKSPPIISINPLFLMRSKFLFISISAVATADLVEFGISRPDSTKFIHQLDPEFYONKGNLSGLKLDLHAPVTGHOVDGKVVNCLGALCEALAKLKLQWCKSKHKKODMLVAGCIQGLGATSAKLKCLDTEBECKIYONCAESEGEVLNDENSRVYVAHSAMLPWIOYNGENRRAEKKLDFSSVRLFFEVFLQKCLKKSIKN"
gene		complement(29879..35351) /gene="Y18D10A.3"
CDS		complement(join(29879..30346,31220..31296,31346..31522,34557..34852,35323..35351)) /gene="Y18D10A.3" /note="predicted using GeneFinder" /codon_start=1 /protein_id="CAA22319.1" /db_xref="GI:3979940"
gene		/translation="MDGSGEKAIGFLQESLNARFWTDKOLATDEPSESSYONVRFRIISERKPSVDKLYKTGEIPARGRGNKYGVSFMSKSMELSLVLTQTPILOOFYKNGKSKTWHTIISKKTVPFISGKLAOIDOELFTYKGFEBOLMELAGIAQOIAAHYKSNAYVLCGPGNNGDGFVCAHRHLOQGFPTSTIYPRKSRMELKSLVQCE TSSIPATIPNLQAPPLIVDALDFGSFHPPTREPTRELTGTVRASGHHVFSIDVPS GMVEIGAPSGNDVDVIPHRSVISTLTPKLCMKNMWGPHFLGGRFVPKSLVDEHELLM

Query Match

gene		POYGFGEQIVKLED" 35614..36138 /gene="Y18D10A.4"
CDS		join(35614..35976,36025..36138) /gene="Y18D10A.4" /note="predicted using GeneFinder" /codon_start=1 /protein_id="CAA22310.1" /db_xref="GI:3979931"
gene		/translation="MNGERQTEAVSKRSNVSRCCGPPCLRLVDQRTLSVCYSDSS HHDGSGYHFFSJDCLSLTSTAIVATAISSSSPPAFVSVSLFMSLVAVYVGVLLVTGMDRRIYVQWFHRLPTRKTIENSFARSSFFECFLKMIIGFWLSP"
CDS		join(39072..39164,40245..40513,41360..41936,44226..44375) /gene="Y18D10A.5"
gene		/note="predicted using GeneFinder: similar to Eukaryotic protein kinase domain; cDNA EST yk307d8.5 comes from this gene; cDNA EST yk4519.5 comes from this gene; cDNA EST yk233b4.3 comes from this gene; cDNA EST yk387f2.5 comes from this gene; cDNA EST yk337g9.5 comes from this gene" /codon_start=1 /protein_id="CAA22311.1" /db_xref="GI:3979932"
CDS		/translation="MNKOLLSCSLKSGQVTVVASVATDGVQDQVEISYDQKVTGN GSRGVFLAKLTSTTNMVAIKRYLDKRRKNRELQIMRNLHNINIKLYKFFYSGSEK KDELIVNLILEYETVYRARRYSKORQOIPHTYKLMYOLKSLALYHSIGTCHR DIKPNLLIDPESGVKLKDFGSAKLVNRPWYSYICSRYYRAPLIFGAINYNSI DWSQATVNAEELLIGQPIFGDSGVQDLEIIVLQTPREQIOSNMNPKERKFPQI KAHPMKRVFVHPPAIDILSKIIETFTSPRTPQAQCHAFDELRLPMDALRPSGR PLPTLMMDGWMGVEGVTSSGDVAGPSA"
gene		47525..53224 /gene="Y18D10A.6"
CDS		join(47525..47593,48112..48208,48610..48751,49616..50045,50501..50818,51553..51971,52106..52364,52847..53224) /gene="Y18D10A.6" /note="similar to Sodium/hydrogen exchanger family" /codon_start=1 /protein_id="CAA22320.1" /db_xref="GI:3979941"
gene		/translation="NMLSVEEQVLRNIEELMFNSLKITLANGVTQWLWESTLNOGTA TSGIMPFALTAISICIFILFQTVQSDSSASVSGAVASDVTVAANKTIVL GEADINATSTEQHGAAYIGVSEKRSALFFILFVIMLALTVHMLIVSKIHMP ESALYALGALLISIIYSRRMSEIALSPDYFVLLPPTIFENAYMLNKGFPSN FVPLIFALFHTTISAMVAGALYILGAGLILEFFEFECFAAIAISVDPVGTAL FQAVKVESLITMLVFGESMLNDVAVSLAATALRAKAPSFNSLIPASEIITSAFVPTLE MFFFSACLGAGIGLISALLFHYDLKRTSLSEFALLILFSYIIPYGAELDLGIMAI LFCGISMSQTRHNVSPIAOITFRHRTISFAETSTPAGYGMARFTIKLNDAPMLI FMSVVLCLGRACNVPLAVLVNOCGRDVOISMKNQIIMFGSMRGAVCFALVLYMDL DKRKSIIITTYVFLITFTTIFLGSAIPISFINRCYNERQRRKRRPRKKESTGN SSALMNSKIQDEMSFSGSDMGFRKSLADTSSAGRTMRDLFRKFTALRELEKDLA ALTKRALASDQMTDSDVDFEGGGVGGGGRMKMDVTPTRGRGSRNNSDVIISAGCG VSGSEHLLIISGSDSSTNEF"
gene		54375..62943 /gene="Y18D10A.7"
CDS		join(44375..54405,54825..54924,54984..55168,55357..55538,56162..56529,56806..57056,57123..57193,57641..57905,58630..58895,58941..59066,59124..59210,59257..59373,59630..59775,61492..61610,61724..61783,61846..61959,62188..62409,62843..62943) /gene="Y18D10A.7" /note="predicted using GeneFinder" /codon_start=1 /protein_id="CAA22312.1" /db_xref="GI:3979933"
gene		/translation="MSQNHIIILPDSFRKSMIITSKIRAPLEOPMNAVAKYCLFVAK YPMPTIIPILITICISMGILNFKIVRGVNTYAPLNTWKEEVAPENNAKDDDH FVGRGLIRROGIYVLNKKDGGSNRQYADDFELKILIMLITNLVKSISAGELITPKD VCHHPNDCEFSNHAFLANLITSKNQNQNSHFNFTIYIYSTATEIDTSLYKGVNVL DHDGHEVNSAWLITVQLKNEKWQLSDEMDFTIQSGEAPSENLNLYRHSATFD

3.9%; Score 68.8; DB 34; Length 152878;

Db	242	TAATTTAAATAATATTTAAATTAATAAATTTTAAATTAATAAATTTAAATTAATAAATTTT	183
Qy		300 gttcttgatgaataatcactactatgtattacagtaatttccatcacatatacatc	359
Db	182	TAATTTAAATAATTTTAAATTAATAAATTTTAAATTAATAAATTTTAAATTAATAAATTTT	123
Qy	360	gttcagcaaaatagctattcagaagctgtaagaagaacagatatcat	408
Db	122	TAATTTAAATAATTTTAAATTAATAAATTTTAAATTTTAAATTAATAAATTTTAAAT	74
RESULT 2			
TD	T78867/C	T78867 standard; DNA; 6152 BP.	
AC	T78867;		
DT	08-OCT-1997	(first entry)	
DE	P. falciparum liver stage antigen-3 genomic sequence.		
KW	Plasmodium falciparum; pre-erythrocyte; liver stage antigen; serum;		
KM	prophylaxis; Thai strain; gene organisation; exon; intron; hydrophobic;		
KM	glycosyl-phosphatidylinositol membrane anchoring sequence; antibody;		
KW	vaccine; immunotherapy; malaria; ds.		
OS	Plasmodium falciparum.		
FT	Key	Location/Qualifiers	
FT	cds	77..5605	
FT		/*tag= a	
FT		/product= liver stage antigen-3	
FT	exon	77..274	
FT		/*tag= b	
FT		/number= 1	
FT	Intron	275..442	
FT		/*tag= c	
FT		/number= 1	
FT	exon	443..5602	
FT		/*tag= d	
FT		/number= 2	
PN	MO9641877-A2.		
PD	27-DEC-1996.		
PR	12-JUN-1996; F00894.		
PR	13-JUN-1995; FR-007007.		
PA	(INSP) INST PASTEUR.		
PR	Daubersies P, Druille P;		
PR	WPI; 97-065464/06.		
PT	Plasmodium falciparum poly:peptide(s) and related nucleic acids -		
PT	derived from the liver stage antigen-3, useful for malaria vaccine		
PT	prodn. and diagnosis		
PS	Claim 20; Fig 1A-E; 69pp; French.		
CC	This sequence corresponds to the genomic sequence encoding a Plasmodium		
CC	falciparum strain K1 pre-erythrocytic liver stage antigen-3 (USA-3)		
CC	protein (M24790). The gene sequence was isolated by screening a		
CC	P. falciparum strain T9/96 library with the serum from a missionary		
CC	treated by prophylaxis (for strain T6/96 see FR9101286). Of 20 clones		
CC	isolated, clone 7295 was used to screen a library generated from Thai		
CC	strain K1. One clone contained a 6.85 kb insert including the sequence		
CC	presented here. The gene organisation comprises a first exon, a short		
CC	168 bp intron and a 5 kb second exon containing a 1.8 kb region encoding		
CC	7 blocks of 4 amino acid repeats and a 3' hydrophobic region		
CC	sequence. The invention relates to new polypeptides of at least 10		
CC	amino acids derived from the USA-3 polypeptide with the exception of		
CC	peptides W24791-4. The USA-3 peptides can be used to raise antibodies		
CC	and as vaccines for immunotherapy of malaria.		
SO	Sequence 6152 BP; 2725 A; 463 C; 1215 G; 1749 T;		

[illegible]

Db	6078	ATTAAACATCCTTTATVTGAAAAA	AAAAAAAAAAAAAAAAAAAAAAAAAGCTTATTTGCATTA	6019
Qy	121	aattgcatatcaccacaatgataacgctaga	-atnaaataggacaagltgataat	179
Db	6018	CACGCAAAATAGCTAAATTTTATCTCTGTAATATATATACATATATATATATATATA	5959	
Qy	180	gacgaataacaacaagaaactcttcacgagcctaagatgaacacattcaaatgatagc	239	
Db	5958	TTATATATATATATATATTTTATATAAACCAAAATATTTAAATATTTACATATATTCGTATGAATA	5899	
Qy	240	atataaagatctactgaataigaataagcttctcgaagcctataaagcagatagaagaa	299	
Db	5898	ATATACATCTTCTTACAGAAATCTATATATTAATTTAAAAACATAACTCTGTAAAAAC	5839	
Qy	300	gtttgacgcaaatatacactactctgatactcagtaattcttcactacactataattc	359	
Db	5838	ATTTATATATTTAAAAATTAATATTTGAAAAATTAATATATGAAAAATTAATAATATTA	5779	
Qy	360	gttagcctaatactagctattagaagatgataagaagaacgaattatatt	410	
Db	5778	ATATATATATATTTTAAAAATTTTAAAAATTTGTAAAAAGATACATATATAT	5728	

	RESULT	3
ID	T72882/C	
AC	T72882 standard; cDNA: 19124 BP.	
DT	T72882;	
DE	12-SEP-1997 (first entry)	
KW	Plasmodium var-7 gene.	
KW	DBL gene family; SABB; sialic acid binding protein; vaccine; therapy;	
KW	Duffy binding like gene; Duffy antigen binding protein; erythrocyte;	
KW	DABP; meirozole; malaria; var-1; var-2; var-3; var-7; immune response;	
KW	Plasmodium; ss.	
OS	Plasmodium vivax.	
FH	Plasmodium falciparum.	
FT	Key	location/Qualifiers
FT	exon	7317..15139
FT		/tag= a
FT		/number= 1
FT	intron	15140..16205
FT		/tag= b
FT		/number= 1
FT	exon	16206..17552
FT		/tag= C
FT		/number= 2
FT		/note= "no stop codon given"
PN	WO9640766-A2.	
PD	19-DEC-1996.	
PF	07-JUN-1996; U09508.	
PR	07-JUN-1995; US-487826.	
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.	
PI	Chittis C, Miller LH, Peterson DS, Sim KL, Su X;	
PI	Wellens TE;	
DR	WPI: 97-052231/05.	
P	P-SDB; W22475.	
PT	New malaria vaccines - contains cysteine-rich DBI family protein	
PT	binding domains homologous domains of the Duffy and sialic acid	
PT	binding proteins	
PS	Claim 4' Page 56-61; 96pp: English.	
CC	This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to the	
CC	Duffy binding like (DBL) family of genes which have homology to the	
CC	Duffy antigen binding protein (DABP) and sialic acid binding protein	
CC	(SABP) conserved regions (see T72889 and T72888 respectively). The var	
CC	family of genes modulate cytoadherence and antigenic variation of	
CC	Plasmodium infected erythrocytes. SABP and the Duffy antigen binding	
CC	protein (DABP) are soluble proteins that appear in the culture	
CC	supernatant after infected erythrocytes release merozoites. DABP and SABBP	
CC	mediate the binding of merozoites and schizonts to the erythrocyte	
CC	surface. These proteins are necessary for erythrocyte invasion by the	
CC	parasite. This sequence can be used in the compositions of the invention.	
CC	The compositions are for the treatment and prevention of malaria, and	
CC	comprise either a nucleotide sequence or encoded polypeptide of the	

Query Match	3.08;	Score 54;	DB 1;	Length 5852;
Best Local Similarity	48.0%;	Pred. No. 0.021;		
Matches 218.	Conservative	0;	Mismatches 230;	Indels 6;
				Gaps 2;

	RESULT	6
062924/C		
ID	062924 standard; cDNA; 8920 BP.	
AC	062924;	
DT	06-DEC-1994 (first entry)	
DE	Cardamoyl-phosphate-synthetase II.	
KW	Cardamoyl-phosphate-synthetase II; CPSII; pscCPSII gene	
KW	malaria; ss.	
OS	Plasmodium falciparum.	
FH	Key	Location/Qualifiers
FT	1226..8401	
FT	cds	
FT	/*tag= a	
FT	/EC_number= 6.3.5.5	
PN	W09412643-A.	
PD	09-JUN-1994.	

Query Match	3.0%;	Score 54;	DB 1;	Length 8920;
Best Local Similarity	47.6%;	Pred. No. 0.022;		
Matches 195;	Conservative	0;	Mismatches 210;	Indels 5;
				Gaps 1;

RESULT 7
ID N90224/c
AC N90224 standard; DNA; 1052 BP.
DT N90224;
DT 1-NOV-1989 (first entry)
DE Malaria-specific DNA insert of clone 41-2.
KW DNA; malaria; clone 41-2; Plasmodium falciparum; protein; vaccine.
OS Plasmodium falciparum
PN EP-322712-A.
PD 05-JUL-1989.
PE 20-DEC-1988; 121299.
PR 30-DEC-1987; DE-831351.
PR (BEHW) Behringwerke.
PA Knapp B, Hundt E, Enders B, Kupper H;
PI Wpf; 89-194071/27.
DR P-PSDB; P90417.
DR New antigenic proteins from Plasmodium falciparum - new encoding nucleic
PT acid sequences and derived antibodies, useful in vaccines, diagnosis etc.
PS Claim 1; Table 17; 25pp; German.

QY 58 aaataagtgaaatgcttcaagaataggaatgagccataaanaattgaggcaatacaagg 117
DB 15550 ATAAATATTAATTTGATAGATAATGAAAAAGATTATATCAAAAAAATTTAAAAA 15609
QY 118 tataatttgatattatcaccaaatgataaactgttagaataaataagacaaagtgtga 177
DB 15610 AAAATTTTATTTAAAAAAAATGATTTAAAAAAAATGAAAAAGAAAGAAAAAAA 15669
QY 178 atgattgataaacaagaactatcatcattgacctgaatgaaacattcaatagat 237
DB 15670 AAACATTTAAAAAAAATATATATATCAAAAAACAAAAAAAATATATTT 15729
QY 238 gcatataaagattactgataaataagcttccagaagctataaagcgatagaag 297
DB 15730 AAAATTAATAATATATATATCAAAAAAATTAATAAATAATGTAATAAAAAATAT 15789
QY 298 aagtttgatgataaataacacttattgtatcaagtaattatccctacatacatat 357
DB 15790 ATACATAAATAATTAATAATTTATTTAATTAATAAATAATTAATAAATTTTAT 15849
QY 358 ttgttagctaaatattgattatagagaatgataagaagaacgatataatgacagt 417
DB 15850 TAAATTAATAAATAATTAATAATTTAATAAATAATTAATAAATAATTAATAAAT 15909
QY 418 ttattatgctcgatgattcctacattattta 452
DB 15910 TAAATGAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 15944

RESULT 11

X20248_02
Continuation (3 of 10) of X20248 from base 200001 (Borrelia burgdorferi polynucleotide s
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248

WP	Fragment Name	Begin	End
WP	X20248_00	1	110000
WP	X20248_01	100001	210000
WP	X20248_02	200001	310000
WP	X20248_03	300001	410000
WP	X20248_04	400001	510000
WP	X20248_05	500001	610000
WP	X20248_06	600001	710000
WP	X20248_07	700001	810000
WP	X20248_08	800001	910000
WP	X20248_09	900001	910715

Query Match 2.9%; Score 50.8; DB 1; Length 110000;

Best Local Similarity 44.8%; Pred. No. 0.16; Mismatches 292; Indels 6; Gaps 1;

QY 20 ttttccacaatgaaatacttcttaattgcaataatgcaataatgtaattggttaccaga 79
DB 40492 TTTTATTTCAAAAATATATATGATCTTGAGATTTTAATACACACGATGATATTTACAAA 40551
QY 80 ataggaaatgagccataaanaattgagcaatatacaaggatataattgtgatatatacca 139
DB 40552 ATAGATTAATAATCTTTAAATAATTTCCCAAGATTTAATGACAACTTTACAAAAC 40611
QY 140 aatgataaactgttagaataaataagcaagttgataatgataataaacaacgaa 199
DB 40612 ATTTAAAAAATAATTAATAAATAATTAATAAATAATTAATAAATAATTAATAAATA 40671
QY 200 ctattcatgagcctaataatgataaactttacaatagatgataaataaagattactga 259
DB 40672 AAAGCTTTGAAAAAGAAAAATACGAGTAATGTAATAAAAAATGGAATATTTGCGCA 40731
QY 260 atgaataagctttcagaagctataaagcgatagaagaagtttgatgaaataatcact 319
DB 40732 ATTACAAAGAAATAGAAATATCTTCACAAAAAGAAATTCATTTTGGGATTTCAAAAT 40791
QY 320 tatgtatatacgaatattatccactataatattgttagctaaataatgattct 379
DB 40792 TCAACTTAAGATAAAT 40845

QY 380 agaagatgataaagaagaacgaatatatatgacgtttattatgtctcgatattct 439
DB 40846 ATGACAAATACGAAAAAAGAGAAATACGCGCTCTTTTAATTTGATTTTTCACACATGAG 40905
QY 440 aacattatttattcctctgatttgagcgaanaattaaacgaatggtlaataalgacg 499
DB 40906 AAAATTTTATCTTACTGCGCTTTCAAGACAACTTAAGCCACGAAATTTATACATCAA 40965
QY 500 cagcatcaatgcaatgctttaaccataaacaattggaagcgaatggtgacatgcaga 559
DB 40966 AAACATTAACAAATAATTAATCTTGACACATCAAAAGATGAAAAAATTTTCAAAATGAAA 41025

RESULT 12

N60472/c
ID N60472 standard; DNA: 4590 BP.
AC N60472;
DT 24-AUG-1991 (first entry)
DE Sequence encoding the ring-infected Erythrocyte Surface Antigen
KW (RESA).
OS Plasmodium falciparum.
FH Key
FT exon Location/Qualifiers
FT exon /tag= a
FT exon 801..995
FT exon 1199..4225
FT exon /*tag= b
FN W08601802-A.
PD 27-MAR-1986.
PE 11-SEP-1985; 006960.
PR 11-SEP-1984; AU-007067.
PR 11-SEP-1984; AU-007066.
PR 10-SEP-1985; AU-047326.
PR (HALL-) HALL INST MED RES.
PI Kemp DJ, Anders R, Coppel RL, Brown G, Saint RB, Cowman AF;
DR WPI; 86-094065/14.
DR P-PSDB: P60569.
PT DNA coding for Plasmodium falciparum antigens - expressing
PT poly:peptide(s) having antigenicity of RESA or FIRA antigens of P
PT falciparum
PS Claim 4; Fig 1; 55pp; English.
CC The inventors claim a novel DNA molecule which comprises a
CC nucleotide sequence corresp. to all or a portion of the base
CC sequence coding RESA (N60472) or FIRA (N60473). RESA and FIRA have
CC antigenicity suitable for providing protective immunity against
CC Plasmodium falciparum malarial infections.
SO Sequence 4590 BP; 1933 A; 437 C; 673 G; 1547 T;

Query Match 2.8%; Score 50.6; DB 1; Length 4590;

Best Local Similarity 46.2%; Pred. No. 0.094; Mismatches 234; Indels 2; Gaps 1;

QY 3 ttt 62
DB 740 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 681
QY 63 agtgaattgtttacagaataggaatgagccataaanaattgagcaatatacaaggtata 122
DB 680 TAAAAATAGTAATTTTATTTTCTGAAAAAATAAATAATTAATTAATTAATTAATTAAT 621
QY 123 ttgtagatatacccaatgataaactgttagaataaataagcaagttgataatgat 182
DB 620 ATATTAAATATATTTTATTTTATGAAAAATGTTATTTATTAATAATTAATTAATAA 561
QY 183 gataatacaacgaacatattcatgagcctaataatgataaactttacaatgataatgata 242
DB 560 AAAAAAATAATATATTAATATTTTATATATAATAA--AATTTTTTAAAAATTTATTTAT 503
QY 243 taaagaattactgaaatgaataagctttccagaagctataaagcgataagaagaagtt 302
DB 502 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAT 443

DB	313	TAIRATG 308
RESULT	15	
ID	003568	
AC	003568	standard; DNA; 6124 BP.
DT	02-AUG-1990	(first entry)
DE	Sequence encoding the SEPA protein of Plasmodium.	
KW	Malaria; SEPA protein; vaccine; ds.	
OS	Plasmodium falciparum.	
FT	Key	Location/Qualifiers
FT	misc-feature	1..6
FT		/*tag= a
FT		/note="Sequence obscured."
FT	misc-feature	121..128
FT		/*tag= b
FT		/note="Sequence obscured."
FT	misc-feature	241..245
FT		/*tag= c
FT		/note="Sequence obscured."
FT	exon	2407..2440
FT		/*tag= d
FT	intron	2441..2598
FT		/*tag= e
FT	exon	2599..3405
FT		/*tag= f
FT	intron	3406..3580
FT		/*tag= g
FT	exon	3581..3721
FT		/*tag= h
FT	intron	3722..3850
FT		/*tag= i
FT	exon	3851..5835
FT		/*tag= j
FT	allele	3087
FT		/*tag= k
FT		/label=Allele II.
FT	allele	3092
FT		/*tag= l
FT		/label=Allele II.
FT	allele	3096
FT		/*tag= m
FT		/label=Allele II.
FT	allele	3140
FT		/*tag= n
FT		/label=Allele II.
FT	allele	3148
FT		/*tag= o
FT		/label=Allele II.
FT	allele	3176
FT		/*tag= p
FT		/label=Allele II.
FT	allele	3182
FT		/*tag= q
FT		/label=Allele II.
FT	allele	3098..3136
FT		/*tag= r
FT		/note="absent in allele II."
FT	misc-binding	3091..3120
FT		/*tag= s
FT		/label=Complementary to Probe A.
FT	misc-binding	3082..3151
FT		/*tag= t
FT		/label=Complementary to Probe B.
FT	misc_signal	/note="Allele II, skips 3098..3136."
FT		485..2526
FT		/*tag= u
FT		/label=Regulatory sequence.
PD	WO9001549-A.	
	22-FEB-1990.	
	10-AUG-1989;	003422.

PR 12-AUG-1988; US-231771.
PA (DART-) Dartmouth College.
PI Inselburg JW, Ezik DJ;
DR WPI; 90-083508/11.
DR P-PSDB; R055526.
PT DNA encoding SERA antigen of Plasmodium falciparum -
PT also DNA encoding signal and regulatory sequences of SERA gene,
PT for diagnosis, prevention and treatment of malaria.
PS Disclosure; P; English.
CC Sequence allows for production of antigenic malarial proteins, useful in
CC diagnosis, prevention and treatment of the disease. The sequence may be
CC used in an expression vector, and signal and regulatory sequences may be
CC used to stimulate production of other products.
SQ Sequence 6124 BP; 2567 A; 602 C; 801 G; 2137 T; 17 Others;

[illegible]

Search completed: April 16, 2000, 04:43:17
Job time: 9896 sec

DB	313	TAIRATG 308
RESULT	15	
ID	003568	
AC	003568	standard; DNA; 6124 BP.
DT	02-AUG-1990	(first entry)
DE	Sequence encoding the SEPA protein of Plasmodium.	
KW	Malaria; SEPA protein; vaccine; ds.	
OS	Plasmodium falciparum.	
FT	Key	Location/Qualifiers
FT	misc-feature	1..6
FT		/*tag= a
FT		/note="Sequence obscured."
FT	misc-feature	121..128
FT		/*tag= b
FT		/note="Sequence obscured."
FT	misc-feature	241..245
FT		/*tag= c
FT		/note="Sequence obscured."
FT	exon	2407..2440
FT		/*tag= d
FT	intron	2441..2598
FT		/*tag= e
FT	exon	2599..3405
FT		/*tag= f
FT	intron	3406..3580
FT		/*tag= g
FT	exon	3581..3721
FT		/*tag= h
FT	intron	3722..3850
FT		/*tag= i
FT	exon	3851..5835
FT		/*tag= j
FT	allele	3087
FT		/*tag= k
FT		/label=Allele II.
FT	allele	3092
FT		/*tag= l
FT		/label=Allele II.
FT	allele	3096
FT		/*tag= m
FT		/label=Allele II.
FT	allele	3140
FT		/*tag= n
FT		/label=Allele II.
FT	allele	3148
FT		/*tag= o
FT		/label=Allele II.
FT	allele	3176
FT		/*tag= p
FT		/label=Allele II.
FT	allele	3182
FT		/*tag= q
FT		/label=Allele II.
FT	allele	3098..3136
FT		/*tag= r
FT		/note="absent in allele II."
FT	misc-binding	3091..3120
FT		/*tag= s
FT		/label=Complementary to Probe A.
FT	misc-binding	3082..3151
FT		/*tag= t
FT		/label=Complementary to Probe B.
FT	misc_signal	/note="Allele II, skips 3098..3136."
FT		485..2526
FT		/*tag= u
FT		/label=Regulatory sequence.
PD	WO9001549-A.	
	22-FEB-1990.	
	10-AUG-1989;	003422.

Oy	2	tcttttcttttcttttctttccaaagaacttcctaataatgcacaaat	61
Db	542	TATGTAAATTATTAATACCTTTATTTATTAATATGTAATTTTAAATACTTTAATTATTAATAA	483
Oy	62	aagcgaaatgcttcaggaataggaatgatggccataaaaatitggcacaataagblata	121
Db	482	TATGTAAATTATTAATACCTTTATTAATATGTAATTTAATAATATGTAATTTAATAACAATT	423
Oy	122	attgtgatattacacaaatagataaacctgttgaaataaataggaaagtligataatga	181
Db	422	TAAATTATTAATAATATGTAATTTAATAACAATTAAATTTAATAATAATATGTAATTTAATAACAATT	363
Oy	182	tgcataatcaacaagaaccttcatatggcctaagtaagaaaccttac--aatagtagc	239
Db	362	TAAATTATTAATAATATGTAATTTAATAACAATTAAATTTAATAATAATATGTAATTTAATAACAATT	303
Oy	240	atataaagaattactgaaataatgaataagcttccagaagctataaaaaaggaatagaagaa	299
Db	302	TAAATTATTAATAATTTAATTTAATAACAATTAAATTTAATAATAATTTAATTTAATAATATTTT	243
Oy	300	gttttgatgaataatcaccttatctgattacagtaaatatttccataactatatactt	359
Db	242	TAAATTATTAATAATTTAATTTAATAATNTTTAATTATTAATAAATAATTAAATTATAAATTTT	183
Oy	360	gttagctaaaatgattatagaagaatgataaagaagaacgaatatacatatgacagcttc	413
Db	182	TAAATTATTAATAATTTAATTTAATAATNTTTAATTTAATAAATAATTAAATTATAAATATTTT	123
Oy	420	attatgtctcgatgcagatccaaactiaattttat	453
Db	122	TAAATTATTAATAATTTAATTTAATAATATTTAAT	89

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

:      TOPOLOGY: linear
:
:      MOLECULE TYPE: DNA (genomic)
:
:      FEATURE:
:          NAME/KEY: misc_feature
:          LOCATION: 3770..4013
:      OTHER INFORMATION: /note="Rp4 origin of DNA transfer (orit) from
:      OTHER INFORMATION: plasmid Rp4"
US-09-056-075-1
```

Query Match	3.28;	Score 57.8;	DB 4;	Length 6243;
Best Local Similarity	47.68;	Pred. No. 0.00022;		
Matches 170;	Conservative 0;	Mismatches 187;	Indels 0;	Gaps 0.

Accession	Sequence	Position
Oy	ttcagaatcgaagaatgtagccataaaatctggcgaatcacagatbatattcgtgcat	132
Oy	73	
Db	TAAATACAGCGCTAAAAAATGAGCGTAAAAATAAAAAATAAAAAATAAAAA	1224
Oy	133 taccacaatgataaactcgtctagataaaatagagcaagcttgataatgataaataa	192
Db	1225 AATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAA	1284
Oy	193 aacgaactatctcctggcctaactgataacatctcaatagatctgatatataaagatca	252
Db	1285 TAAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAATAT	1344
Oy	253 ctgaaatgataaagctcttcagaagctataaagcagtagaagagcttgcgtgaata	312
Db	1345 AAAATATATAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAATAT	1404
Oy	313 taccactatctgataaacgtaactcttccctacactataatattgctgaataat	372
Db	1405 TATTTTATATTTAAAGTTGAAAAAAATTTTATATATATATAAATCTTTGAAGAAAGA	1466
Oy	373 agctatctgaagagctgataagaagaacgaatataigcgcagcttctatctgct	429
Db	1465 ATATATAAAATGAGCCCTTTATATAAGCCCAATTTTTTCTATATAGCTAATATAGAGCT	1521

1 RESULT 3
 2 US-08-487-826B-13/C
 3 Sequence 13 Application US/08487826B
 4 Patent No. 5993827
 5
 6 GENERAL INFORMATION:
 7 APPLICANT: Sim, Kim L.
 8 APPLICANT: Chitnis, Chetan
 9 APPLICANT: Miller, Louis H.
 10 APPLICANT: Peterson, David S.
 11 APPLICANT: Su, Xin-zhan
 12 APPLICANT: Wellens, Thomas E.
 13 TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 14 TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 15 NUMBER OF SEQUENCES: 45
 16
 17 CORRESPONDENCE ADDRESS:
 18 ADDRESSEE: Knobb Martens Olson & Bear
 19 STREET: 620 Newport Center Drive 16th Floor
 20
 21 CITY: Newport Beach
 22 STATE: California
 23 COUNTRY: US
 24 ZIP: 92660
 25
 26 COMPUTER READABLE FORM:
 27 MEDIUM TYPE: Floppy disk
 28 COMPUTER: IBM PC compatible
 29 OPERATING SYSTEM: PC-DOS/MS-DOS
 30 SOFTWARE: PatentIn Release #1.0, Version #1.25
 31
 32 CURRENT APPLICATION DATA:
 33 APPLICATION NUMBER: US/08/487,826B
 34 FILING DATE: 10-SEP-1993
 35 CLASSIFICATION: 435
 36 ATTORNEY/AGENT INFORMATION:
 37 NAME: Israelsen, Ned
 38 REGISTRATION NUMBER: 29,655
 39 REFERENCE/DOCKET NUMBER: NIH121.001CP1


```
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,105
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 2504 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-484-105-15
```

```
Query Match          2.6%; Score 47; DB 1; Length 2504;
Best Local Similarity 45.5%; Pred. No. 0.044;
Matches 95; Conservative 17; Mismatches 97; Indels 0; Gaps 0;

QY 121 aattgcatattatccacaatgataaactgltagaataaataagacaagltgataatg 180
DB 2190 AACTTGCGTGAGTACCAATAAGCTATCTTTGGATGAATAAGCAATTTACCGATTG 2249
QY 181 atgataataacaagaactatcaltgacctgaatgataagaacatttacaatgatalga 240
DB 2250 AACCAAGATATAAACACTTCCCTTACCTTACCTTGAAATTTTAAATGTTTACTCTTAAT 2309
QY 241 tataaaagattactgaataatgaataagctttcagaagcctataaaagcgaatagaagaag 300
DB 2310 TATAAATTACGACTTAATTAATCTTTTAATTTGCCCATGAMAAARARMRARMRMRW 2369
QY 301 tttagatgaataatcacttattgtacta 329
DB 2370 RMMWMAWMMMGATACTACTACTCTTATTA 2398
```

```
RESULT 13
US-08-484-106-15
Sequence 15, Application US/08484106
Patent No. 5614618
GENERAL INFORMATION:
APPLICANT: STILLMAN, Bruce
APPLICANT: BELL, Stephen P
APPLICANT: KOBAYASHI, Ryuji
APPLICANT: ROSE, Jasper
APPLICANT: FINE, Margit
APPLICANT: MCNALLY, Francis J
APPLICANT: LAURENSEN, Patricia
APPLICANT: HERSKOWITZ, Ira
APPLICANT: LI, Joachim J
APPLICANT: GAVIN, Kimberly
TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESS: FLEHR, HORBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
```

```
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,106
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 2504 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-484-106-15
```

```
Query Match          2.6%; Score 47; DB 1; Length 2504;
Best Local Similarity 45.5%; Pred. No. 0.044;
Matches 95; Conservative 17; Mismatches 97; Indels 0; Gaps 0;

QY 121 aattgcatattatccacaatgataaactgltagaataaataagacaagltgataatg 180
DB 2190 AACTTGCGTGAGTACCAATAAGCTATCTTTGGATGAATAAGCAATTTACCGATTG 2249
QY 181 atgataataacaagaactatcaltgacctgaatgataagaacatttacaatgatalga 240
DB 2250 AACCAAGATATAAACACTTCCCTTACCTTACCTTGAAATTTTAAATGTTTACTCTTAAT 2309
QY 241 tataaaagattactgaataatgaataagctttcagaagcctataaaagcgaatagaagaag 300
DB 2310 TATAAATTACGACTTAATTAATCTTTTAATTTGCCCATGAMAAARARMRARMRMRW 2369
QY 301 tttagatgaataatcacttattgtacta 329
DB 2370 RMMWMAWMMMGATACTACTACTCTTATTA 2398
```

```
RESULT 14
US-08-323-170B-1
Sequence 1, Application US/08323170B
Patent No. 5733772
GENERAL INFORMATION:
APPLICANT: Williamson, Kim C.
APPLICANT: Kaslow, David C.
TITLE OF INVENTION: Cloning and Expression of Plasmodium
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
```


Search completed: April 16, 2000, 04:36:19
Job time: 9753 sec

RESULT	2	AA701731/c	EST	19-DEC-1997
LOCUS	AA701731/c			
DEFINITION	AA701731	628 bp	mRNA	EST
ACCESSION	AA701731			
VERSION	AA701731.1	GI:2704931		
KEYWORDS	Onchocerca volvulus.			
SOURCE	Onchocerca volvulus.			
ORGANISM	Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Filarioidae; Onchocercidae; Onchocerca.			
REFERENCE	1 (bases 1 to 628)			
AUTHORS	Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.			
TITLE	Genes expressed in molting L3 larvae of Onchocerca volvulus			
JOURNAL	Unpublished (1997)			
COMMENT	On Sep 19, 1997 this sequence version replaced gi:1517220.			
	Contact: Steven A. Williams			
	Molecular Parasitology			
	Smith College Department of Biological Sciences			
	Department of Biological Sciences, Clark Science Center, Smith			
	College, Northampton, MA, 01063, USA			
	Tel: 4135853626			
	Fax: 4135853786			
	Email: genome@smith.edu			
	Seq primer: Bluescript SK.			
FEATURES	Location/Qualifiers			
source	1..628			
	/organism="Onchocerca volvulus"			
	/strain="Kumba, Camerouns"			
	/db_xref="taxon:6282"			
	/clone="SMOV3MCA1144"			
	(clone_id="Onchocerca volvulus molting L3 larva cDNA			
	(SL69MLW-OvmL3)"			
	/db_stage="molting L3"			
	/lab_host="XLI-Blue MRF"			
	/note="vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:			
	Xho I; Filarial nematode parasite of humans. Third-stage			
	larvae, L3, were isolated from infected black flies in			
	Cameroun (forest strain). The L3 were cultured in 20% FCS			
	in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in			
	culture. L3 of O. volvulus molt to fourth-stage larvae by			
	day 5 in culture. mRNA was isolated from approximately			
	6000 molting larvae (ML3), 2000 larvae from day 1, 2 or 3			
	in culture, and converted to double-stranded cDNA using			
	reverse transcriptase and oligo(dT) followed by RNase H			
	and DNA pol I. The library was constructed in the lambda			
	Uni-Zap XR vector and has 1 x 10E6 independent			
	recombinants and the average insert size is ~1200 bp. The			
	library was constructed by Sara Lustigman and Michelle			
	Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams			
	The library is available from Dr. Sara Lustigman (email:			
	slustigman@bc.org)."			
BASE COUNT	181 a 121 c 125 g 201 t			
ORIGIN				
Query Match	21.0%;	Score 373.8;	DB 37;	Length 628;
Best Local Similarity	79.9%;	Pred. No. 4.3e-65;		
Matches 453;	Conservative 0;	Mismatches 112;	Indels 2;	Gaps 1;
Oy	1056	gattccagaatcgatgcattatgataa--aacgtgctgaccaaagtatgaacaattga	1113	
Db	622	ggtttcagatgatacactgtccatttatgrraaacactggttgcacaaatgtatgacaaatttga	563	
Oy	1114	actgtgtgacgcggttgtgtccaccatccaaattcataacgcatatctgcatcgggaca	1173	

Db	562	ACAGGTTGTCCGGATGGTCCACCATCAAGAAATTTTCATATCCGGCATACAGCATTCGGTACA	503
Oy	1174	attgagattgtaaaagcaggttgatgtgatttcagatacctcaatcgtgacctaactgtttaa	1233
Db	502	ATTGTGTGGCAATGCAATGGTTGATTCGTAATTCGGAAACTTCGAAGTTGAGCCGTAACTGCTCTTA	443
Oy	1234	tcagcttcacgtataaagcgtcttaccggtatgcagatagcataacttggtaacaaat	1293
Db	442	TCAGCTTCACATGTAAGAAGCATTTGATTCGATATGCTCGATCAACTTTTGTCCAAACANT	383
Oy	1294	gagatgaacggaatgacaaacagttgtgtttacaacaaataccacgtlgaatcagagatcgt	1353
Db	382	GGATGGAATGAGATGAGACAAACACTGATGTGACAAAATAACACAGTGGATTTAAACAGAGCT	323
Oy	1354	gtacgcgcacaactatgcatlgaatcaaatggaagtgaattccggcgaactlgaagctccact	1413
Db	322	GTACCGTGTCAATTTACGAGGAATTAACAGGAAGTTCAATTCGGCTTACCGTAGCTCCACT	263
Oy	1414	tcacatcattagcgcaactctctlgtacaataaagacacttcacataacatgctcttcgat	1473
Db	262	TCATCATTTTGGGCAATCTCGCTGATGCTATTAACCTTTTCACGATACATGCTCTTCAAT	203
Oy	1474	gcatcagctgltataaaattgatgtgtatgttaagattgtgtccaatcatttcgctgctca	1533
Db	202	GGATTACAGAGTAATTAAGTTGACAGAGTAATGATGTTGGACACATTCATTTCTGTTCA	143
Oy	1534	ctcttcgaacacctatgttcaaccggatctggaataagacaatgcaataagtgtaagaaga	1593
Db	142	CTCTTCAACCGTTTATTAACCGGAATAGCTTTTACGAGGTGATAATAGCAATACGCTT	83
Oy	1594	gcaataagacgcatacattcattccagaa	1620
Db	82	ATTAACGAGCAATGATATATACGTA	56
RESULT	3		
LOCUS	A1322117/c		
DEFINITION	A1322117 639 bp mRNA	EST	22-DEC-1998
ACCESSION	SMOV3MCAM12G08SK	Onchocerca volvulus molting L3 larva cDNA	
VERSION	(S166MLM-OvML3)	Onchocerca volvulus cDNA clone SMOV3MCAM12G08 5',	
KEYWORDS	A1322117.1	GI:4056268	
SOURCE	EST.		
ORGANISM	Onchocerca volvulus.		
REFERENCE	Onchocerca volvulus.		
AUTHORS	Eukaryotes; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;		
TITLE	Flarioidae; Onchocercidae; Onchocerca.		
JOURNAL	1 (bases 1 to 639)		
COMMENT	Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.		
	Genes expressed in molting L3 larvae of Onchocerca volvulus		
	Unpublished (1997)		
	On Jan 14, 1998 this sequence version replaced gi.1797212.		
	Contact: Steven A. Williams		
	Molecular Parasitology		
	Smith College Department of Biological Sciences		
	Department of Biological Sciences, Clark Science Center, Smith		
	College, Northampton, MA, 01063, USA		
	Tel: 4135853826		
	Fax: 4135853786		
	Email: genome@smith.edu		
FEATURES	Seq primer: pbluescript SK.		
Source	Location/Qualifiers		
	1..639		
	/organism="Onchocerca volvulus"		
	/strain="Kumba, Cameroons"		
	/db_xref="taxon:6282"		
	/clone="SMOV3MCAM12G08"		
	/clone_1ib="Onchocerca volvulus molting L3 larva cDNA		
	(S166MLM-OvML3)"		
	/dev_stage="molting L3"		
	/lab_host="X11-Blue MRF"		
	/note="Vector: Lambda uni-ZAP XR; Site 1: Eco RI; Site 2:		

Accession	Version	Keywords	Source	Organism	Reference	Authors	Title	Journal	Comment
AY1291	1		AA668071/c	LOCUS	610 bp	EST	20-NOV-1997		
AY1291	1		AA668071	DEFINITION	SMOV3MCAM02C01SK Onchocerca volvulus molting L3 larva CDNA (SLI96MLM-Ovml3)				
AY1291	1		AA668071	ACCESSION	AA668071				
AY1291	1		AA668071.1	VERSION	GI:2629570				
AY1291	1			KEYWORDS	Onchocerca volvulus.				
AY1291	1			SOURCE	Onchocerca volvulus				
AY1291	1			ORGANISM	Onchocerca volvulus				
AY1291	1			REFERENCE	Onchocerca volvulus				
AY1291	1			AUTHORS	Eukaryotes; Metazoa; Nematoda; Secernentea; Spirurida; Splurida; Filarioidea; Onchocercidae; Onchocerca.				
AY1291	1			TITLE	1 (bases 1 to 610)				
AY1291	1			JOURNAL	Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.				
AY1291	1			COMMENT	Genes expressed in molting L3 larvae of Onchocerca volvulus				
AY1291	1				On Oct 30, 1997 this sequence version replaced gi:2160807.				
AY1291	1				Contact: Steven A. Williams				
AY1291	1				Molecular Parasitology				
AY1291	1				Smith College Department of Biological Sciences				
AY1291	1				Department of Biological Sciences, Clark Science Center, Smith				
AY1291	1				College, Northampton, MA, 01063, USA				
AY1291	1				Tel: 4135853826				
AY1291	1				Fax: 4135853786				
AY1291	1				Email: genome@smith.edu				
AY1291	1				Seq primer: PBLinescript SK.				
AY1291	1				Location/Qualifiers				
AY1291	1				1. 610				
AY1291	1				/organism="Onchocerca volvulus"				
AY1291	1				/strain="Kumba, Camerouns"				
AY1291	1				/db_xref="taxon:6282"				
AY1291	1				/clone="SMOV3MCAM02C01"				
AY1291	1				/clone_id="Onchocerca volvulus molting L3 larva CDNA (SLI96MLM-Ovml3)"				
AY1291	1				/dev_stage="molting L3"				
AY1291	1				/lab_host="XLI-Blue MRP"				
AY1291	1				/note="Vector: Lambda uni-ZAP XR; Site:1: Eco RI; Site:2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (Forest strain). The L3 were cultured in 20% FCS in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volvulus molt to fourth-stage larvae 1 day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and Oligo(dT) followed by RNase H				

and DNA pol I. The library was constructed in the `lambda` Uni-zap XR vector and has 1 x 10E6 independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: `slustig@nybc.org`)."

BASE COUNT	187 a	106 c	138 g	178 t	1 others
------------	-------	-------	-------	-------	----------

	Query Match Similarity	74.0%	Score 355.8:	DB 37;	Length 610;	
	Best Local Similarity	20.0%;	Pred. 1.7e-61;			
	Matches	450;	Conservative	0;	Mismatches 155;	Indels 0; Gaps 0;
OY	847	aatactgcatcgtgcatttgatagaagaacgtgtgatccgcatatttgtatagctgaagt	906			
Dd	609	ATACAAATTCCTGCATGAGTAATAATCTCAGATCATCGACTAAGCTATTTTAAACATGAGCT	550			
OY	907	tcttggccaagccattaataactcgttttgatatatcccgaatttatagcaaatattatccaaga	966			
Dd	549	TCTTGGCCACGCCATTATGAGATNTATGATATTCCAAATGTGTAAGTAATAATCACTTGCCACA	490			
OY	967	gcacatccatcacgaccttgaaatttcocacagatcacocgtttcacatccatcgacaagcag	1026			
Dd	489	GCACACACCTCTTCTTATTCATTCAGATCTCTACCGTATCACCTTTCCGTCATCCACAGTACAA	430			
OY	1027	gaatggacaacccgcgcagaaaglatcaaacgittlcaagaatcgcatgttcattatgataa	1086			
Dd	429	GAAATCCACTAAATGACACAGAATATATCTGATGTTCTGATGATCACAAAGTCCATTGTGATAT	370			
OY	1087	actgctgaccacatgatatggcaattgaaactgtgttgaccogtltgtgtccaccatccaaaatt	1146			
Dd	369	ACGTATCTCCGATCATTTGCTCATATGGAATAGGTGTACCGATGGTGTCCACCTTCCAGAAATC	310			
OY	1147	tcaatacggcatactcgtgcacgcgcgaacatttgatttgaanaacagtttgatttccaagt	1206			
Dd	309	TGCTATTCGACACACAGGCGATTGTGTACCACTTGTGTCTGGAATGCGTGTGCTATTCGGAT	250			
OY	1207	acctcaatctgtgcactaaactgttllatlcagcttccatgtlaaagaatgtactcgatat	1266			
Dd	249	ACTTCAAGTTCGCACTAACGCTTTGTGACGCTTCATATGAAAACACACTGTATACGATAT	190			
OY	1267	gaacgatcaactcttgatlaacaatalatgataibaacgaagaigaacaagtlgttttaaa	1326			
Dd	189	GCTCGGTCTATTTTTGTATATGAAATTTCCGATTAATAATGAATATACATTTATGTTGATTA	130			
OY	1327	aaatatccacagtgatctcagagatcgtgtgaacgcgaacaactgtcatgaatcaaagtgaat	1386			
Dd	129	AAGACACCTCGAGGATTTTAAGGACACCGCTGATGTTCAACATTATCAATGAATTCATCGAACT	70			
OY	1387	gaaatccggcaactctgacgctccaccttcaatccatlaacgagcaacctcttgatcataaaga	1446			
Dd	69	TCAATTTCCACGCTACCCTGCGCTCCACACATCTGAAACGACATTCATTCGGATCGTAAAC	10			
OY	1447	cctttcac 1454				
Dd	9	CCTTTTAC 2				
RESULT	6	AA585626	776 bp	mRNA	EST	30-DEC-1997
LOCUS	AA585626/c					
DEFINITION	SMDD9CA349SK Brugia malayi L3 molting-day 9 larva cDNA					
	(SMD9M7ML-BmlJ3d3)					
	Brugia malayi CDNA clone SMD9CA349 5', mRNA					
	sequence.					
ACCESSION	AA585626					
VERSION	AA585626.1					
KEYWORDS	EST.					
SOURCE	Brugia malayi.					
ORGANISM	Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spiruri-					
	d; Onchocercidae; Brugia.					

REFERENCE
1 (bases 1 to 615)
AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
TITLE Genes expressed in molting I/3 larvae of *Onchocerca volvulus*
JOURNAL Unpublished (1997)
COMMENT Contact: Steven A. Williams

FEATURES	Location/Qualifiers
source	1. .615

BASE COUNT	172 a	124 c	123 g	189 t	7 others
ORIGIN					

Query Match	16.9%;	Score 301.2;	DB 44;	Length 615;
Best Local Similarity	77.6%;	Pred. No. 1.3e-50;		
Matches 389;	Conservative	0;	Mismatches 109;	Indels 3;
				Gaps 2

[illegible]

Qy 1480 cgttataaattgttatttgaagtgtgcacatcaattctcggccacctcg 1539

Db 138 CGAGTATTAAAGTTGACAGATTAATTGATGTTGGACCCATTCATTTTCGTCCACTTCT 139

Qy 1540 acaccattgttaaccgggaatcgaataagacaatgcataagtgtagtacagaagcaata 1599

Db 138 ACACCGTTATACCCGGAATAGCATTTTACAGCAGTGATTAATAGCCCTTACGCTTTATTAC 79

Qy 1600 agacgaatcatcatcttcagaa 1620

Db 78 CGAGCATGATTAATATACGTA 58

RESULT	10
AA901444/c	
LOCUS	AA901444
DEFINITION	AA901444 531 bp mRNA EST 26-MAR-1999
	SMOVC3MCA03B05 Onchocerca volvulus molting L3 larva cDNA

ACCESSION	AA901444
VERSION	AA901444.1
KEYWORDS	GI:3037198
SOURCE	EST. <i>Onchocerca volvulus</i> .

REFERENCE
AUTHORS

1 (bases 1 to 531)
Williams, S.A., Liczote-Waniewski, M., Laney, S., Lustigman, S.,
Haller, L., Allen, M., Bowles, L., Geisel, S., Jost, S., Kucaba, T.,
Martin, J., Steptoe, M., Theising, B., White, Y., Wyllie, T.,
Chappell, J., Person, B., Gibbons, M., Harvey, N., Pape, D.,
Chamberlain, A., Morales, R., Schurr, R., Rittler, E., Kohn, S.,
Underwood, K., and Marra, M.

Email: genome@smith.edu
The library was constructed by Sara Lustigman and Michelle Litowicz-Waniewski in the Laboratory of Dr. S.A. Williams. The library is available from Dr. Sara Lustigman email slustigman@nybc.org When requesting this clone from Dr. Lustigman please reference the Williams lab clone id - SMOVCAM03B05
Seq primer: -40ml3 fwd. Rev. ET from Amersham
High quality sequence stop: 484.

FEATURES

SOURCE

NOTE: Vector: lambda Uni-ZAP XR; Site1: Eco RI; Site2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (Forest strain). The L3 were cultured in 20% FCC in IMDM+NCIC 135 and collected after day 1, 2, or 3 in culture. L3 of *O. volvulus* molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (mL3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda

KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus
Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Filarioidea; Onchocercidae; Onchocerca.

REFERENCE 1 (bases 1 to 466)
AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL Unpublished (1997)
COMMENT On May 9, 1995 this sequence version replaced gi:802407.
Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK.
Location/Qualifiers

FEATURES
source
1..466
/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMML3C0758"
/clone_id="Onchocerca volvulus molting L3 larva cDNA (SL96MLW-OvML3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (forest strain). The L3 were cultured in 20% FCS in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volvulus molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (ML3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda Uni-Zap XR vector and has 1 x 10E6 independent recombinants and the average insert size is ~1300 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustig@nybc.org)."

BASE COUNT 126 a 91 c 93 g 144 t 12 others
ORIGIN

Query Match 15.1%; Score 269.4; DB 36; Length 466;
Best Local Similarity 75.3%; Pred. No. 3e-44;
Matches 327; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

OY 1187 aagcagttgatttcagatcctcaatcgtgcactactgtttatcaagcttcacatgt 1246
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 466 ATGCAATTGATTCGGAAGCGTCAAGTGTAGCGCTGCTTATCACCTTCATGT 407
OY 1247 aaagcattgactcagatcagatcattgtgtacaaataatgatgaagaa 1306
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 406 AAGAGCCTTGTATTCGATATCTCGATCACTTTTGTCAACANCAATGAGNATGGA 347
OY 1307 tgacaacagttgtgttcaaaataaccagtgatgcagatcgtgtacgcgaacat 1366
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 346 TGACAACAACATGATGTGACAACATACACGTGANNNTAAGAACGTGTACGTGTACAT 287
OY 1367 tgcatagaatcaaatggaattccgcgaacttgacgctcacctcatcatcagac 1426
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 286 TACAGAAATCAANGGAAGGTCAATTCGCGNACCTGACGTCACTTCATCATTTTCGCG 227
OY 1427 aacctcttgatcaaaagaccttcaataaactgaccttgatgacatgacatgacat 1486
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 226 ATCTGTGCTGATGATTAAGCTTTACGTANANATGCTTCAATGATTACGATAT 167

OY 1487 taaattgattgttaagtgtccacattcattcgtcgcaccttcgaacacat 1546
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 166 TAACTTACAGTAATTTATGTTGACACACATTCATATTTCGTACACTTCTACACCGT 107
OY 1547 tgcacccggaatcgaataagaacatgcaataagtgtgtacagaagaacataagacga 1606
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 106 TATCAACCGGATATACACTTACGACGATGATTAAGCAATACGCTTATTAACCGAAGCA 47
OY 1607 tcaatcttcagaa 1620
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 46 TGATATATATCACTA 33

RESULT 13
AA841200/c
LOCUS
DEFINITION MB3D6AAAG10T3 Brugia malayi day 6 post-infection third stage larvae
SAM96MLW-BmlJ3d6 Brugia malayi cDNA clone 3D6AAAG10 5', mRNA
sequence.
AA841200
AA841200.1 GI:2922536
VERSION
KEYWORDS
SOURCE
ORGANISM
Brugia malayi.
Brugia malayi.
Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Filarioidea; Onchocercidae; Brugia.
REFERENCE 1 (bases 1 to 354)
AUTHORS Blaxter, M.L., Waterfall, M., Daub, J., Lizotte-Waniewski, M., Baron, L. and Jones, S.J.
TITLE Genes expressed in day six post-infection, third stage larvae of Brugia malayi
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:228052.
Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JF, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
The ABI trace of this sequence can be viewed at
http://www.sanger.ac.uk/brugia/3D6/MB3D6AAAG10T3.html
Seq primer: T3.

FEATURES
source
1..354
/organism="Brugia malayi"
/strain="T3S Labs"
/db_xref="taxon:6279"
/clone="3D6AAAG10"
/clone_id="Brugia malayi day 6 post-infection third stage larvae SAM96MLW-BmlJ3d6"
/sex="mixed"
/dev_stage="third stage larvae, six days after infection"
/lab_host="E. coli XLI-Blue"
/note="Vector: lambdaZapRII (UniZap XR); Site_1: Eco RI (5' end); Site_2: Xho I (3' end); Brugia malayi is a lymphatic filarial nematode parasite of humans. mRNA was prepared from third stage larvae of Brugia malayi isolated from the peritoneal cavity of jirds six days after infection. The mRNA was converted to double stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNApol I. The library has 2 x 10E5 independent recombinants and average insert size was 900 base pairs. The library was constructed by Michelle Lizotte-Waniewski. The library is available from The Filarial Genome Project Resource Center: contact Dr. S.A. Williams, Clark Science Center, Smith College, Northampton, MA 01063 USA phone +1 413 585 3826 fax +1 413 585 3786 email genome@smith.edu."

BASE COUNT 105 a 81 c 84 g 84 t

Email: genome@smith.edu
Seq primer: Bluescript SK.
Location/Qualifiers
1. 395

/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SmlJ3C0318"
/clone.lib="Onchocerca volvulus molting L3 larva cDNA
(SL96MLW-Ovml3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-Zap XR; Site-1: Eco RI; Site-2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNA pol I. The library was constructed in the lambda
Uni-Zap XR vector and has 1 x 10E6 independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Waniewski in the laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustig@nycb.org)."

BASE COUNT 118 a 76 c 88 g 112 t 1 others
ORIGIN

Query Match 14.6% Score 260.4; DB 36; Length 395;
Best Local Similarity 84.0%; Pred. No. 1.9e-42;
Matches 305; Conservative 0; Mismatches 37; Indels 1; Gaps 1;
QY 748 gcaagtttctgctgcygcaacacctgtttaacagctccgaatcccttgctgtgaacat 807
Db 375 GCAGCTGCAGCACACACATTTGCGANCAACTGCTCCAAATCCTTGCTGCACAT 316
QY 808 tctgtcgaacacatctgctatcttgctccttaagttaactgatactgcatgtag 867
Db 315 TGTGTCGTGCACATTCACATGTTGTTCTTAATTGTAATTAATTGACATTTGTTAG 256
QY 868 aaaagctgtagatccgcatatctgtatagctgctcttgccagccatlaactc 927
Db 255 AACAGTTGTGATCGATCCGCATATTTAGAGCTGAGCTTCTTGCCAGCCATTAAATCA 196
QY 928 gttagatatccaaattattagcaaatattatcaagagacacatccatcattaga 987
Db 195 GTTGATATTCCAAGTGTGAGCAAAATATTGTCNAAGCACACACATCAAGATTAAAT 136
QY 988 atttccacagatcacgcttaacatcatcgacaagaaggaatgacacaacgcgcagaaa 1047
Db 135 AATTCCACCTTATCACCGTATACATCAACGAAGCATGAATGTACAAACACACAGCAAT 76
QY 1048 gtatcaaggtttcagaatcgatgcatcattatgataaactgctggaaccaatgataca 1107
Db 75 GTATCAACTGTTTCAGATTCACATGTCATTTATGTGTACACTGG-TGACCAATGTGATGCA 17
QY 1108 aat 1110
Db 16 AAT 14

Search completed: April 16, 2000, 03:09:09
Job time: 6193 sec


```

563 nTrpGlyGluGlnThrsSerSerAsnValProArgProGlyGlnPheArgI 580
292 .....ACAACTGTTGTCATTTGCTTTCATTCATTTATTTGTTACC 333
      |||:|||||: ||| |||:
580 leProHisSerLeuThrMetValProAspGlnGlyGlnLeuCysValAla 596
      |||:|||||: |||:|||||: |||: |||:
334 AAGTTGATGTCATATGAGTACATGCTTTTACATGAGACCTGATAA 383
      |||:|||||: |||:|||||: |||: |||:
597 AspArgGluAsnGly...ArgIleGlnCysPheHisAlaGlnThrGlyAs 612
      |||:|||||: |||:|||||: |||: |||:
384 AACAGTTAGTGCACAGATTGAGTATCTGAATC...ACAATGCTTTTC 430
      |||:|||||: |||:|||||: |||: |||:
612 nPheValLys...GlnIleLysHisGlnGlnPheGlyArgGluValPhea 628
      |||:|||||: |||:|||||: |||: |||:
431 AACCTCAAAATTCCTCCGATCCAGTATCCGTTATGATAATTTGGATGCT 480
      |||:|||||: |||:|||||: |||: |||:
628 ValAlaSerTyraLPro.....Gly 634
      |||:|||||: |||:|||||: |||: |||:
481 GGACCAACCGGTCAACCACTTCAATTGCTATCATTTGTCAGCAGTTTA 530
      ||| |||:|||||: |||:|||||: |||
635 Gly.....ValLeuTyraLAlaValAlaSnGlyLysProTyTy 646
      |||:|||||: |||:|||||: |||: |||:
531 TCATTAATGACATGCGGATTCTGAACCGTTGATCTTCTGCGCGTTG 580
      ||| |||: |||: |||: |||: |||:
646 r.....GlyTySerAlaProV 652
      |||:|||||: |||:|||||: |||: |||:
581 TCATCTCTGCTTTGTCGATGATGTAACGTCATCTGAGAAATTTCTA 630
      |||:|||||: |||:|||||: |||: |||:
652 AlGlnIlePheMetLeuAsnPhSerAsnGlyAsP..... 663
      |||:|||||: |||:|||||: |||: |||:
631 AATGCTGATGATGCTCTTGTGTAATATATTG.....CTAATATATT 674
      ||| |||: |||: |||: |||: |||:
664 .....IleLeuAspThrPheIleProAlaArgLysAsnPh 675
      |||:|||||: |||:|||||: |||: |||:
675 GGAATATCCACAGATTTAATGCTGGCCAGAAGCTCAGCTTACAAAT 724
      |||:|||||: |||:|||||: |||: |||:
675 eAspMetProHisAspIleAlaAlaAspAspGlyThrValTyValG 692
      |||:|||||: |||:|||||: |||: |||:
725 ATGCGATGATCAGCAGCTTTCTATCAATGCCAGATGCT..... 765
      |||:|||||: |||:|||||: |||: |||:
692 LysAspAlaHisAlaAsnAlaValAlaTrpLysPheSerProSerLysAlaGlu 708
      |||:|||||: |||:|||||: |||: |||:
766 ...ATTACATTAAAGAACCAAT.....AGCGAATGTGTGCACC 803
      |||:|||||: |||:|||||: |||: |||:
709 HisArgSerValLysLysAlaGlyIleGluValGlnLutLeuThrGln 725
      |||:|||||: |||:|||||: |||: |||:
804 ACAATGTTCAAGAACCAAGGATTGCGAGCTGTTAAACAGGTGTGCCG 853
      |||:|||||: |||:|||||: |||: |||:
725 rGluIlePheGlnThrHisIleArgSerArgProLysThrAsnGlnSerV 742
      |||:|||||: |||:|||||: |||: |||:
854 CACCAAAACCTGCTGCAGCTGGCAGCACTTCGTTTACTCAAGAAAGATCT 903
      |||:|||||: |||:|||||: |||: |||:
742 AlGlnLysGlnThrGlnGlnLysGlnAsnValValGlnLutLeuAsnAl 936
      |||:|||||: |||:|||||: |||: |||:
904 GCA.....GAACGAGAAATATCATT...GATGTACGAC 935
      ||| |||: |||: |||: |||: |||:
757 AlaGlyValAlaSerThrGlnGlnLysGlnAsnValValGlnLutLeuAsnAl 773
      |||:|||||: |||:|||||: |||: |||:
936 TGTATCAACACCCCTTGAA.....ATTACGATGATTAATCAAGCTT 976
      ||| |||: |||: |||: |||: |||:
773 aGlyValProThrGlnGlnLysGlnAsnValValGlnLutLeuSerAlaG 790
      |||:|||||: |||:|||||: |||: |||:
977 TGCACATTGATTACCGTCACCGCTTCGACTTCGACATATGAGCAACCT 1026
      |||:|||||: |||:|||||: |||: |||:
790 LysAlaSerThrGlnGlnLysGlnSerValValGlnLutLeuSerAlaGly 806
      |||:|||||: |||:|||||: |||: |||:
1027 GTAATACTGCTGCAGTACAAAATGGAATCTGCATCTCACCATTGGCTT 1076
      ||| |||: |||: |||: |||: |||:
807 ValSerThrGlnGlnLysGlnSerValValGlnLutLeuSerAlaGly 823
      |||:|||||: |||:|||||: |||: |||:
1077 CTCATGTTTATGGTTTAAAGCATTTGCATTGATT...GCTGCCGTATTA 1123
      |||:|||||: |||:|||||: |||: |||:

```

```

823 lSerPheValLeuIleIleThrLeuLeuIleProIleAlaValLeuI 840
1124 TTACCAATTGCTTAATTTGCT 1146
      |||:|||||: |||: |||:
840 leAlaIleAlaIlePheIleArg 847
      |||:|||||: |||: |||:
seq_name: A_Geneseq_36:P94856

```

```

seq_documentation_block:
ID P94856 standard; protein; 693 AA.
AC P94856.
DT 27-JUN-1990 (first entry)
DE Expression plasmid pUCPIC1799 BglII gene product.
KW alpha-amidating; pAX799; alpha amide; ds.
OS Synthetic.
PN EP-299790-A.
PD 18-JAN-1989.
PF 15-JUL-1988; 306508.
PR 17-JUL-1987; JP-177184.
PR 05-DEC-1987; JP-306867.
PA (SUNR) Suntory Ltd.
PI Ohsuye K, Kifano K, Tanaka S, Matsuo H, Mizuno K;
DR WPI: 89-017279/03.
DR N-PSDB: N94527.
PT Recombinant C-terminal alpha amidating enzymes of Xenopus laevis -
PT and their precursors deoxyribonucleic acid encoding sequences.
PS Disclosure: 7pp: English.
CC The sequence encodes a derivative of the mature C-terminal alpha-
CC amidating enzyme from plasmid pAX799.
CC The plasmid was screened from an E.coli library using plasmid pXA457
CC to screen a larger library.
CC Although pXA799 is similar to pXA457 at the N-terminus, it has an area of
CC hydrophobic elements suggesting a membrane function.
CC See also N93060
SQ Sequence 693 AA:

```

```

alignment_scores:
Quality: 94.50 Length: 232
Ratio: 0.945 Gaps: 12
Percent Similarity: 43.103 Percent Identity: 21.552

```

```

alignment_block:
US-09-323-427-3 x P94856 ..

```

```

Align seg 1/1 to: P94856 from: 1 to: 693

```

```

97 TGTGACCACTTCAATTAACATCAATTTTAATACAGTAAATGATTCGA 146
      ||| |||:|||||: |||: ||| |||
491 CysGlnProThrAspValAlaValAspProIleThrGlyAsnPhePheVa 507
      |||:|||||: |||:|||||: |||: |||:
147 AGGACATGTTTATGTGAAGAAGCTTTATGATCAAGAAGTGGCCGTAAG 196
      ||| |||
507 lAlaAspGlyTyT..... 511
      |||:|||||: |||:|||||: |||: |||:
197 ATGAGGTGAGCGTCAAGTTGCCGAATTTCACTTCATTGATTCATGC 246
      ||| |||
512 .....Cys 512
      |||:|||||: |||:|||||: |||: |||:
247 AATGTTCCGCTACAGATCTCTGAATCCAGTGGTATTTTGTGA..... 291
      ||| |||: |||: |||: |||: |||:
513 Asn...SerArgIleMetGlnPheSerProAsnGlyMetPheIleMetG 528
      |||:|||||: |||:|||||: |||: |||:
291 ..... 291
      |||:|||||: |||:|||||: |||: |||:
528 nTrpGlyGluGlnThrsSerSerAsnLeuProArgProGlyGlnPheArgI 545
      |||:|||||: |||:|||||: |||: |||:
292 .....ACAACTGTTGTCATTTGCTTTCATTCATTTATTTGTTACC 333
      ||| |||: |||: |||: |||: |||:
545 leProHisSerLeuThrMetIleSerAspGlnGlyGlnLeuCysValAla 561
      |||:|||||: |||:|||||: |||: |||:
334 AAGTTGATGTCATGAGTACATGCTTTTACATGGAAGCTGATAA 383
      |||:|||||: |||:|||||: |||: |||:

```



```

826 SerSerpheAlaIleCysAsnThrGlyIuIleLysTyrValAsnValn 842
459 CGGTATGAAATTTGGATGT..... 480
      ||| ||| ||| |||
842 rHsValAlaGluIleValAspSerIleGlyValIleLysProValSerT 859
481 ..GGACCAACCGGTCAACAGTTCAATTTGGCTATTCAGTGGTCAAGTT 528
      ||| ||| ||| ||| ||| ||| |||
859 hGlyAsnIleSerIleProLysAsnPhetHrValAlaValAlaGlnaGlu 875
529 TATCAATAATGACATGCATCTGTGAACCGTACTTCTGCCGGT 578
      ||| ||| ||| ||| ||| ||| ||| ||| |||
876 TyrIleGlnIleGlnValLysProValValAsp.....CysAlaThr 890
579 TGTTCATTCCTGCTTTTGTGATGATGATGATGATGATGATGATGATTC 628
      ||| ||| ||| ||| ||| ||| ||| ||| |||
890 r.....TyrValCysAsnGlyAsnThrHisCysLeuLysLeuL 903
629 TA.....AATGCT.....GAT 639
      ||| ||| ||| ||| ||| ||| ||| ||| |||
903 euThrGlnIlyThrSerAlaCysGlnThrIleGlnAsnAlaLeuAsnLeu 919
640 GGATGCGCTTGATTAATATTTGCTAAATTAATTTGGAATATCCACACAGA 689
      ||| ||| ||| ||| ||| ||| ||| ||| |||
920 GlyAlaArgLeuGlnSerLeuMetLeuAsnAspMet..... 931
690 TTTAATGCGTGCACAGAGCTCAGCTATACAAATTCGGGATCATCAG 739
931 ..... 931
740 ACCTTTTCTATCAATGCGATCATGATTATTAAGACCAATATAC 789
      ||| ||| ||| ||| ||| ||| ||| ||| |||
932 .....IleThrValSerAspArgLysLeu 939
790 GAATGCTTCACACACATGTCAGACACCAAGATTCGGAGCTTTAA 839
      ||| ||| ||| ||| ||| ||| ||| ||| |||
940 GluLeuAlaThrValGlnArgPheAsnAlaThrAlaLeuGlyLysLeu 956
840 AACAGGTGGTCCGCGAGCAAAACCTGCTGAGTCCGCAACTTCGTTAC 889
      ||| ||| ||| ||| ||| ||| ||| ||| |||
956 sleuGlyLeuGlyLeuTyrPheAspGlyLeuSerSerLeuLeuProProLysI 973
890 TCAAGAAAGATCTCGACAGACCGGAGATATCATTTGAT.....GTACGA 933
      ||| ||| ||| ||| ||| ||| ||| ||| |||
973 leGlyLysArgSerAlaValGlnAspLeuLeuPheAsnLysValValThr 989
934 ACTGATATACACACCTTGAATTAGCGATGATATACAGCTTTGCCAGT 983
      ||| ||| ||| ||| ||| ||| ||| ||| |||
990 SerGlyLeuGlyThrValAspAspAspTyrLysLysCysSerSerGlyTh 1006
984 TGATTTACCTCACCGCTGCTGTCGACATATGAGCAACCTGTATATAC 1033
      ||| ||| ||| ||| ||| ||| ||| ||| |||
1006 rAspValAlaAspLeuValCysAlaGlnTyrTyrAsnGlyIleMetValL 1023
1034 TTGCTGCACTACAAATGGAATTCGATCCACATTTGGCTTCTCAATG 1083
      ||| ||| ||| ||| ||| ||| ||| ||| |||
1023 euProGlyValValAspGlyAsnLysMetSerMetTyrThrAlaSerLeu 1039
1084 TTTATGGGTTTAAAGCAT...GCATTTGATTTGCTGCGCTCATTTATTCAT 1130
      ||| ||| ||| ||| ||| ||| ||| ||| |||
1040 IleGlyGlyMetAlaLeuGlySerIleThrSerAlaValAlaVal.ProP 1056
1131 TT.....CGTTAAATTTGCTCCAAATACAGA 1156
      ||| ||| ||| ||| ||| ||| ||| ||| |||
1056 heAlaMetGlnValGlnAlaArgLeuAsnTyrValAlaLeuGln 1070

seq_name: A_Geneseq_36:W88310
seq_documentation_block:
ID W88310 standard; Protein; 374 AA.
AC W88310:
DE 26-APR-1999 (first entry)
E. coli colitose or glucose transferase.

```

```

KW O antigen; O111 antigen; wbdm gene; colitose transferase;
KW glucose transferase; diarrhoea; haemorrhagic colitis; diagnosis.
OS Escherichia coli.
PN W09850531-A1.
PD 12-NOV-1998.
PE 01-MAY-1998; A00315.
PR 22-JUL-1997; A0-008162.
PR 01-MAY-1997; A0-006545.
PA (UNSY ) UNIV SYDNEY.
PI Reeves PR, Wang L;
PI WPI: 99-059669/05.
DR N-PSDB; X06748.
DR Nucleic acid molecules specific for bacterial polysaccharide
PT antigens - useful for detecting specific strains in, e.g. food,
PT faeces or patient samples
PS Disclosure: Fig 7; 165pp; English.
CC This is the amino acid sequence of the protein encoded by the wbdm
CC gene of a gene cluster (see X06748) involved in the biosynthesis of
CC the Escherichia coli O111 antigen. The protein shows high
CC homology with TSE of Yersinia enterocolitica, and is predicted to
CC be a colitose or glucose transferase. The use of nucleic acid
CC molecules derived from particular assembly and transport genes,
CC particularly wbd (transferase), wzx (flippase) and wzy (polymerase)
CC genes, within O antigen gene clusters improves the specificity of
CC methods for the detection and identification of O antigens, e.g. in
CC testing food- or faecal-derived samples, or samples from patients.
CC The O antigen is a major virulence factor of enteropathogenic E.
CC coli strains that cause diarrhoea and haemorrhagic colitis.
SQ Sequence 374 AA;

```

```

alignment_scores:
  Quality: 89.50      Length: 363
  Ratio: 0.533      Gaps: 18
  Percent Similarity: 46.281      Percent Identity: 21.763

```

alignment_block:

US-09-323-427-3 x W88310

Align seg 1/1 to: W88310 from: 1 to: 374

```

62 TTGACAAATGCTGTGCAAGGTGACGCCGAAATTTGATGTGACCAACTTCA 111
      ||| ||| ||| ||| ||| ||| ||| ||| |||
38 LeuThrGlyIleSerGlnValLysProThrGlnAsnIleAsn..... 51
112 ATACAAATCAATTTTAATACAGTAATGCATTCGAAAGGAC.....A 152
      ||| ||| ||| ||| ||| ||| ||| ||| |||
52 .....IleHisTyrValAsnMetAspLysAsnPhaArgS 63
153 TGTATTGTGAAAGCTTTTATGATCAAGAGGTGCGTAAATGATGAG 202
      ||| ||| ||| ||| ||| ||| ||| ||| |||
63 erPheheArgAlaLeuPheGlnValLysLysIleLeuAlaLeuLys 79
203 GTGACGTCAAGTTGCCGAATTTCACTTCATTTGATTCATGCAATGTT 252
      ||| ||| ||| ||| ||| ||| ||| ||| |||
80 ProAspIle.....IleHisSerHisMetPheHisAlaAsn.. 91
253 GCGCGTACACAGATCTGTAATCCAGTGGTATTTTGTACACACACTGT 302
      ||| ||| ||| ||| ||| ||| ||| ||| |||
92 .....I 92
303 TGTCAATTTGCTTTCATCAATTTATTTGTACCAAGTTG...ATCGCAT 349
      ||| ||| ||| ||| ||| ||| ||| ||| |||
92 IePheSerArgPheIleArgMetLeuIleProAlaValProLeuIleCys 108
350 ATCGAAGTCAATGCTTTTACATGGAAGCTGATAAAGCACTAGTGCACAG 399
      ||| ||| ||| ||| ||| ||| ||| ||| |||
109 ThrAlaHisAsnLysAsnGlnGlyGlyAsnAlaArgMetPheCysTyrXr 125
400 ATTGAGATATCTGCAATTCACAACTGCTTTTCAACATCAATATGTCGGAT 449
      ||| ||| ||| ||| ||| ||| ||| ||| |||
125 glu.....SerAspPheLeuAlaSerIleThrHisAsnValS 138

```

```

450 GCCAGTATGCCGTTATGAATTTGGATGGTGAACCAACCGGTCAACAG 499
    :: |||||:::|||||::: |||||:::
138 erlysgluValaValaGluInleuhe.....lleAlaArglySalatThrPro 152
    |||::: |||::: |||||:::
500 TTCAATTTGCTATCATTGGTCACCGCAGTTTATCAT..... 534
    |||::: |||::: |||||:::
153 LysAsnLyslleValaGluIleProAsnPhelAsnThrAsnLysPheAs 169
535 .....AAATGACATCGCATTCCTGGAACCGTTGATA 565
169 pheAsnIleAsnValaArgLysLysThrArgAspAlaPheAsnLysAs 186
566 CTTTCGCGCGGTGTCATTCCTGCTTTGTCGATGATGTAACGGTGAT 615
    |||||::: |||:::
186 spserThrAlaValaLleuValaVala.....GlyArg 196
616 ACTGTGGA.....ATTCTAATGCTATGATGCTGC 647
    |||||::: |||||:::
197 LeuValaGluAlaLysAspTyrProAsnLleuAsn.....Al 209
648 TCTTGATTAATATTGCTAATAATTGGAATATCCACAGAT..... 690
    |||||::: |||||:::
209 aileAsnHisLleuIleLeuSerLysThrSerAsnCysAsnAspPheIleL 226
691 ..TTAATGCGTGGCAAGAAGCTCAGGTATACAATATGCGGATCGATCA 738
    |||||::: |||||:::
226 euleuIlealagLyspGlyAlaLleuArgAsnLysLleuAsp..... 240
739 CAGCTTTTCTATCATGCCAGATCAGTATTACCATTA..... 777
    |||||::: |||||:::
241 .....LeuValaCysGlnLleuAsnLleuValaAspLysValaPhePheLe 254
778 .GAACCAATAGCGAATGTGTGACACCAATGT..... 810
    |||||::: |||||:::
254 uGlyGlnArgSerAspIleLysGluLleuMetCysAlaAlaAspLeuPheY 271
811 .....TCGAACCAACAAGATTCGGA...GCTGTTAAACAGGTGCT 849
    |||||::: |||||:::
271 alLeuSerSerGluTyrpGluGlyPheGlyLeuValaIaGluAlaMet 287
850 GCCGCAGCAAAACCTGCTGCAGCTGCGCAACTGCTTTACTCAAGAAAG 899
    |||||::: |||||:::
288 AlaCysGluTyrProValaValaIaThrAspSerGlyGlyValaLysGluVa 304
    ::: |||||::: |||||:::
900 ATCTGAGAACCGAGAAATATCATGTATGATGACAGACGATATC..... 942
    ::: |||||::: |||||:::
304 lValaGlyProHisAsnAspValaIleProValaSerAsnHisLleuLeuA 321
943 .....AACCCCTTGAATTAAGCGATGATATCAAGCTTTG 978
    |||||::: |||||:::
321 laGluLysIleAlaGluThrLleuLysIle...AspAsnAlaIaArgLys 336
979 CCAGTTGATTTAGTCACCGTCGACTTCGCAACATTAAT 1017
    ::: |||||::: |||||:::
337 lleIleGlyMetLysAsnArgGlyTyrIleValaSerAsn 349
seq_name: A_Geneseq_36:W13825
seq_documentation_block:
ID W13825 standard; Protein: 1226 AA.
AC W13825;
DT 04-JUN-1997 (first entry)
DE Yeast transcription regulatory factor SRB8.
KW Transcription regulatory factor; suppressor of RNA polymerase B;
KW SRB8; RNA polymerase II; holoenzyme; SWI/SNF.
OS Saccharomycetes cerevisiae.
PN W09708301-A1.
PD 06-MAR-1997.
PE 28-AUG-1996; U14192.
PR 31-AUG-1995; US-521872.
PR 11-OCT-1995; US-540804.
PR 26-JAN-1996; US-590399.
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.

```

```

PI Chao DM, Koleske AJ, Thompson CM, Young RA;
DR WPI: 97-179258/16.
DR N-PSDB: T59908.
PT Purified RNA polymerase II holoenzyme - comprises RNA polymerase II
PT and one or more regulatory proteins or SWI/SNF proteins
PT polymerase B proteins or SWI/SNF proteins
PS Claim 11; Fig 10a-b; 14pp; English.
CC Novel yeast SRB (suppressor of RNA polymerase B) proteins SRB4, SRB5,
CC SRB6, SRB7, SRB8, SRB9, SRB10 and SRB11 (W13821-28) are transcription
CC regulatory factors that act as positive and negative regulators of
CC RNA polymerase II activity, and are components of the RNA polymerase
CC II holoenzyme. They were identified using methods designed to
CC identify transcription factors involved in RNA polymerase II
CC C-terminal domain (CTD) function. SRB8 and SRB9 appear to repress
CC CTD activity. Genomic clones (T59904-11) for the SRBs have been obtd.
CC SRBs can be used to treat diseases resulting from alteration or
CC deletion of the SRB gene, pref. by gene transfer technology. They
CC can also be used in in vitro transcription of DNA and to identify
CC cpts. that modify gene transcription.
SQ Sequence 1226 AA;

alignment_scores:
    Quality: 88.50      Length: 391
    Ratio: 0.468      Gaps: 16
    Percent Similarity: 48.338      Percent Identity: 20.716

alignment_block:
US-09-323-427-3 x W13825 ..

Align seg 1/1 to: W13825 from: 1 to: 1226

14 TTATGCTTTCGTACTACATTAATGATTCATTCGATTCGCGTT 63
   ::::: |||||::: |||||::: |||||:::
358 MetMetProSerLeuTyrArgLeuAsnIleLleuIleThrTyrGlyI 374
64 GACAAATGGTGTGCAAGTGAGCGCAAAATTAATGTGACCAATTCAT 113
   ::: |||||::: |||||:::
374 erleLysValaProThrTyrIleArgLysLleuIle..... 385
114 AACAAATCAATTTTAATACAGTAATGATTCGAAAGACATGTTAATGTA 163
   |||||::: |||||::: |||||::: |||||:::
386 ..SerSerLysLeuLeuTyrLeuGlnAspSerAsnAspLysPheValHis 401
164 AAGCTTTATGATCAAG...AAGTTGCCGTAATGATCAAGGTGAGCT 210
   ::::: |||||::: |||||::: |||||:::
402 ValGlnLeuLeuIleAsnLysIleSerProLeuMetLys..... 415
211 CAAGTTGCCGAATTTCACTTCATTTGATTCATGCAATGTTGCCGCTAC 260
   ::::: |||||::: |||||::: |||||:::
416 .....SerGlnTyrAsnMetValaLleuArgAsnValaMetGluT 428
261 AGCATCTCGAATCCAGCTGATTT..... 286
   ::|||::: |||||::: |||||:::
428 yAspValaLysPheTyrGluIlePheAsnPheAspIleValaGluIle 444
287 .....TTGTAACACACACGTTGTCATTCGTTTCATTCGATTAAT 327
   ::: |||||::: |||||::: |||||:::
445 ThrGluGlnIleLysMetArgIleLeuSerAsnAspIleThrAsnLeuCl 461
328 GTTACCAAGTTGATCGTGCATATGATGATCAATGCTTTTACATGGAAC 377
   |||||::: |||||::: |||||::: |||||:::
461 neuserLysThrProLeuSerLleLysIleMetValaIaGluTyrPtyL 478
378 TG.....ATAAACACTTTAGTCGACAGATTTGAGGAT 409
   |||||::: |||||::: |||||::: |||||:::
478 euSerHisLeuCysSerGlyIleLeuSerSerValaAsnArgThrValleu 494
410 CTGAATACACACTGCTTTTC.....AAATCAATATGCTCCGATGCA 453
   |||||::: |||||::: |||||::: |||||:::
495 lleLysIlePheLysIlePheCysIleAspLeuGluValaPheHisHisPh 511
454 GTATGCCGTTATGAATTTTGGATGGTGGACCAACCGGTCAACAGTTCA 503

```

```

511 ephelystripileglupheillevaltyrhiscinleuLeuSeraspilec 528
504 ATPTGCTA.....TCATTGCTC 520
528 IuserleuglualaleuemetaspileleuLeucystyrilglnlytleuphe 544
521 AGCCATTATCATTAATGACATGCGATTCTG...AACCGTTGATA.. 565
545 SerGlnphe...IleasnaspHisIleleuPheThrlysthrhelleph 560
566 .....CTTTCTGCGCGGTTTCATTCCTCTGTTTCGATGA 602
560 eilletyrlyslsvalleuylsleuysaspvalproalatyasnvaltr 577
603 TGGTACGGGATACGTGGAATTCATAATGCTGATGATGCTGCTTG 652
577 hserPheMetProPheTrpLysPheMetLysasnPheProPheval 593
653 ATTAATATTGCTTAATAATTTGGAATATCCACAGATTTAATGCTGCG 702
594 LeuLysValaspasnspLeuArgIleGlnleuGlnSerValtyrAsnAs 610
703 CAAGAAGCTCAGCTATACAAATATGCGGATCGATCACAGCTTTTCATCA 752
610 pgluLysleuLysThrGlnLysleuLysasnspLysSerGlnValleu 627
753 ATGCAGATCATGATTAACATTAAGACCA..... 784
627 ysValtyrSerMetIleasnasnSerAsnGlnIleValGlnIntrTP 643
785 ...ATAGCAATGCTGTTCGACCAATGTTTCAG.....AACCAACAAGA 825
644 AsnPheProGlnValPheGlnValAsnIleArgPheleuLeuHis...As 659
826 TTCGGAGCTTTAAACAGGTGTCGCCGACAA..... 859
659 nserGlnIleIleAspThrAsnThrSerLysGlnPheGlnLysAlaArg 676
860 .....AACCTGCTGCAGCTGCGCAGCTTC 883
676 snAsnValMetLeuLeuIleAlaThrAsnleuysGlnTyrrAsnLysPhe 692
884 GTT.....TACTCAAGAAAGATCTGCAGAACCGGAGATA..... 919
693 MetSerIlePheleuLysArgLysAspPheThrAsnLysasnleuIleG 709
920 .....TCATTGATGATACGACTGATATCAACACCCCTTGA 953
709 nleuIleSerLeuLysleuLeuThrPheGlnValIthrGlnAsnValleuG 726
954 AATTACGATGATATCAAGCTT 976
726 lylleuGlnTyrrIleIleArgleu 733

seq_name: A_Geneseq_36:R06400
seq_documentation_block:
ID R06400 standard; protein; 776 AA.
AC R06400.
DT 17-DEC-1990 (first entry)
DE Lambda gtl0ch2 encoded C-terminal alpha amidating enzyme.
KW C-terminal alpha amidating enzyme; human thyroid gland;
KW lambda gtl0ch2.
OS Homo sapiens.
PN M09008190-A.
PE 26-JUL-1990.
PR 17-JAN-1990; J00042.
PA (SUNR) SUNTORI LTD.
PI Ohsume K, Kitano K, Tanaka S;
DR WPI:90-254034/33.
DR N-PSDB: Q05631.

```

```

PT C-terminal alpha amidating enzymes from human thyroid gland -
PT converting C-terminal glycol peptide(s) or protein, and new DNA
PT coding sequences.
PS Disclosure; pp; English.
CC cDNA libraries were prepared from human thyroid gland poly(A) RNA
CC in lambda gtl1 and gtl10. These were screened using probes derived
CC from DNA encoding similar enzymes in Xenopus laevis. Three clones
CC were isolated from the gtl10 library, including gtl0ch2.
CC Restriction analysis indicated that gtl0ch2 encoded a different
CC type of enzyme to one of the other two clones. The DNA can be
CC inserted into vectors for expression in E.coli or (more efficiently)
CC in animal cells.
CC The enzyme is useful for prodn. of physiologically active alpha-
CC amidated peptides from their C-glycol precursors.
CC See also Q05630 and Q05632.
S0 Sequence 776 AA;

alignment_scores:
Quality: 87.00 Length: 322
Ratio: 0.592 Gaps: 13
Percent Similarity: 45.652 Percent Identity: 18.944

alignment_block:
US-09-323-427-3 x R06400
Align seg 1/1 to: R06400 from: 1 to: 776

97 TGTGCGCACTTCATTAACATCAATTTTAATACAGTAATGATTCGA 146
111 |||||.....:
434 CysGlnProThrAspValAlaValAspProGlyThr..... 445
147 AGCAATGTTTATGTAAGGCTTTATGATCAAGAAGTGGCGTAAG 196
111 |||||.....:
446 GlnAlaIleTyrrValSerAspGlyTyr..... 454
197 ATCAAGGTGAGCGTCAAGTTGCCGAATTTCACTTCATTTGATTCATGC 246
455 .....Cys 455
247 AATGTTGCCGCTACAGATCTCGAATCAGCTGATTTTGTACAAAC 296
111 |||||.....:
456 Asn...SerArgIleValGlnPheSerProSerGlyLysPheIleThrG 471
297 A.....ACNG 301
471 nTrpGlnGlnIleuSerSerGlySerSerProLeuProGlnGlnPheThrV 488
302 TT.....GTCAATTCGTTTCATCCATTAATT.....GTTACCAAGTT 339
111 |||||.....:
488 alProHisSerLeuAlaLeuValProLeuLeuGlnLysIleuLysValAla 504
340 GATCGTGCATAT...CGACTACATGCTTTTACATGGAAGCTGATTAAC 386
111 |||||.....:
505 AspArgGlnAsnGlnArgGlnLysPheLysThrAspThrLysGlnPhe 521
387 AGTACTGCACAGATGAGTATGTAATCAATCAACAACTGCTTTCAACAG 436
111 |||||.....:
521 eValArgGlnIleLysHisSerSerPheGlnArgAsnValPheAlaIleS 538
437 AATATGTCGCCGATGCCAGTATGCGGTTATGAATTTGATGGTGGACCA 486
111 |||||.....:
538 eTyrIlePro..... 541
487 ACCGGTCAACAGTTCAATTTGCTATCATTTGTCAGCCAGTTATCATTA 536
111 |||||.....:
542 .....GlyLeuPhePheAlaValAsnGlnLysProHisPheLys 554
537 ATGACATGCGATTCGAAACGTTGATCTTCTGCGCGGTTGTCATAT 586
111 |||||.....:
555 .....AspGlnGlnProValGlnGlnLysPhe..... 562
587 CCTGCTTTGTCATGATGTAACGGTGAATCATCTGTGAATTTCTAAATGCT 636

```



```

726 ValGluThrIysMetGluAsnLysProThrSerSerGluLeuGlnLysM 743
887 TACTCAGAAAGATCT.....GCAGACCCGAGAAATATCATGATGTA 930
743 eGlnGlnLysGlnLysLeuIleLysGlnProGlySerGlyValProVal 759
931 CGAAGTATATCAACACCTTGAAATAGAGATGATATCAAGCTTTGGC 980
760 .ValLeuIleThrThrLeuLeuValIleProValValValLeuAlaI 776
981 AGTTGATTACGTC 994
776 LeAlaIlePheIle 780

```

seq_name: A_Geneseq_36:R06399

```

seq_documentation_block:
ID R06399 standard; protein; 974 AA.
AC R06399;
DT 17-DEC-1990 (first entry)
DE Lambda gt10ch201 encoded C-terminal alpha amidating enzyme.
KW C-terminal alpha amidating enzyme; human thyroid gland;
KW Lambda gt10ch201.
OS Homo sapiens.
PN WO9008190-A.
PD 26-JUL-1990.
PR 17-JAN-1990; J00042.
PR 17-JAN-1989; JP-005878.
PA (SONR ) SONTORY LTD.
PI Ohshuye K, Kitano K, Tanaka S;
DR WPI; 90-254034/33.
DR N-PSDB; Q05630.
PT C-terminal alpha amidating enzymes from human thyroid gland -
PT converting C-terminal glycol peptide(s) or protein, and new DNA
PT coding sequences.
PS Disclosure: JP; English.
CC cDNA libraries were prepared from human thyroid gland poly(A) RNA
CC in lambda gt11 and gt10. These were screened using probes derived
CC from DNA encoding similar enzymes in Xenopus laevis. Three clones
CC were isolated from the gt10 library, including gt10ch201.
CC Restriction analysis indicated that gt10ch201 encodes a different
CC type of enzyme to the other two clones. The DNA can be inserted
CC into vectors for expression in E.coli or (more efficiently) in
CC animal cells.
CC This enzyme is useful for prodn. of physiologically active alpha-
CC amidated peptides from their C-glycol precursors.
CC See also Q05631 and Q05632.
CC Sequence 974 AA;

```

alignment_scores:

Quality:	87.00	Length:	322
Ratio:	0.592	Gaps:	13
Percent Similarity:	45.652	Percent Identity:	18.944

alignment_block:

US-09-323-427-3 x R06399

Align seg 1/1 to: R06399 from: 1 to: 974

```

97 TGTGACCACTTCATTAACATCAATTTTAATACAGTAATGCATTGCA 146
111 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
631 CysGlnProThrAspValAlaValAspProGlyThr.....642
147 AGCAGCATGTTTATGTGAAGGCTTTATGATCAAGAAGTGGCGTAANG 196
111 |||||:|||||:|||||:|||||:|||||:|||||:|||||
643 .GlyAlaIleTyrValSerAspGlyTyr.....651
197 ATGAAGGTGAGCTCAAGTTGCCGAAATTCACCTTCATTGATTGATCG 246
652 .....Cys 652

```

```

247 AATGTCCCGGTACAGATCTGTGAATCCAGCGTGATTTTGTAAACAC 296
111 |||||:|||||:|||||:|||||:|||||:|||||:|||||
653 Asn...SerArgIleValGlnPheSerProSerGlyLysPheIleIle 668
297 A.....ACTG 301
668 nPrGluGluIleSerSerGlySerSerProLeuProGluInPheThrV 685
302 TT.....GTCAATTCGTTTCATCCATTATTT.....GTAAACAAAGTT 339
111 |||||:|||||:|||||:|||||:|||||:|||||:|||||
685 aLProHisSerLeuAlaLeuValProLeuLeuGlyGlnLeuCysValAla 701
340 GATCGTCATAT...CGACTACAATGCTTTTACATGACAGTGAATAAAC 386
111 |||||:|||||:|||||:|||||:|||||:|||||:|||||
702 AspArgGluAsnGlyArgIleGlnCysPheLysThrAspThrLysGluH 718
387 AGTTAGTGCACAGATTGACGTATGTGAATATCAACACTGCTTTTCAACT 436
111 |||||:|||||:|||||:|||||:|||||:|||||:|||||
718 eValArgGluIleLysHisSerSerPheGlyArgAsnValPheAlaIle 735
437 AATTTGCCCGAATGCCAGTATGCCGTTATGAATTTTGGATGGTGACCA 486
111 |||||:|||||:|||||:|||||:|||||:|||||:|||||
735 eTyrIlePro.....738
487 ACCGTCACACAGTTCAATTTGCTATCATTTGTCACAGCTTATATCA 536
111 |||||:|||||:|||||:|||||:|||||:|||||:|||||
739 .....GlyLeuLeuPheAlaValAsnGlyLysProHisPheLys 751
537 ATGACATCGCATTTGAAACCGTTGATCTTCTGCGCGGTTCGCAT 586
111 |||||:|||||:|||||:|||||:|||||:|||||:|||||
752 .....AspGlnIleProValGlnGlyPhe.....759
587 CCGCTTTGTCATGATGTAACCGTATACGTGTGAATTTGAAATGCT 636
111 |||||:|||||:|||||:|||||:|||||:|||||:|||||
760 ....ValMetAsnPheSerAsnGlyIleIleLeuSplePhe.....772
637 GATGATGTGCTCTTGTAATTTGCTAATTAATTTGGAATATCAAC 686
111 |||||:|||||:|||||:|||||:|||||:|||||:|||||
773 .....LysProValArgLysHisPheAspMetProIle 783
687 AGATTAAATGCTGCCAAGAACCTGACSTATACAAATGTCGGATGAT 736
111 |||||:|||||:|||||:|||||:|||||:|||||:|||||
783 sAspIleValAlaSerGlnAspGlyThrValLysIleGlyAspAlaHis 800
737 CACAGCTTTTATCATCAATGCAATGATGATTAACATTAAGAACCAAT 786
111 |||||:|||||:|||||:|||||:|||||:|||||:|||||
800 HisAsnThrValIlePheLysPheThrLeuThrGluLysLeuHisAspSer 816
787 AGCGAATGTGTCAGCACCATGTTTCAGAACCAAGAGATTGAGACTGT 836
111 |||||:|||||:|||||:|||||:|||||:|||||:|||||
817 ValLysLysAlaGlyIleGluValGlnGluIleLysGluAlaGluAla 833
837 TAAACAGCTGTGCGCCAGCAAAACCTGCTGCAGCTGCGCAACTTCGTT 886
111 |||||:|||||:|||||:|||||:|||||:|||||:|||||
833 ValGluThrIysMetGluAsnLysProThrSerSerGluLeuGlnLys 850
887 TACTCAGAAAGATCT.....GCAGACCCGAGAAATATCATGATGTA 930
111 |||||:|||||:|||||:|||||:|||||:|||||:|||||
850 eGlnGlnLysGlnLysLeuIleLysGlnProGlySerGlyValProVal 866
931 CGAAGTATATCAACACCTTGAAATAGAGATGATATCAAGCTTTGGC 980
111 |||||:|||||:|||||:|||||:|||||:|||||:|||||
867 .ValLeuIleThrThrLeuLeuValIleProValValValLeuAlaI 883
981 AGTTGATTACGTC 994
111 |||||:|||||:|||||:|||||:|||||:|||||:|||||
883 LeAlaIlePheIle 887

```

seq_name: A_Geneseq_36:R74171

seq_documentation_block:

ID R74171 standard; Protein; 3038 AA.
AC R74171;

```

DT 18-JAN-1996 (first entry)
DE Aspergillus terreus triol polyketide synthase.
KM Triol polyketide synthase; TPKS; HMG-CoA reductase inhibitor;
KW hypercholesterolemia; LDL- cholesterol.
OS Aspergillus terreus.
FH Key Location/Qualifiers
FT region 181
FT region /label= keto-acyl synthase motif
FT region 654..658
FT region /label= acetyl/malonyl transferase motif
FT region 985..994
FT region /label= dehydratase motif
FT region 1446..1450
FT region /label= methyl transferase motif
FT region 1932..1937
FT region /label= enoyl reductase motif
FT region 2164..2169
FT region /label= keto reductase motif
FT region 2498
FT region /label= acyl carrier protein motif
FT misc_difference 282..288
FT misc_difference /label= misc feature
FT misc_difference 1450..1460
FT misc_difference /label= misc feature
FT misc_difference 1603..1612
FT misc_difference /label= misc feature
FT misc_difference 2521..2535
FT misc_difference /label= misc feature
PN M09512661-A.
PD 11-MAY-1995.
PR 28-OCT-1994; U12423.
PR 02-NOV-1993; US-148132.
PA (MERI ) MERCK & CO INC.
PI Conder MJ, Davis CR, Hendrickson LE, Mcada PC, Rambosek J,
PI Reeves CD, VincI VA;
PI MPI: 95-193816/25.
DR N-PSDB: Q92323.
PT Novel DNA encoding triol poly-ketide synthase - used to isolate and
PT identify homologues of triol poly-ketide synthase, and in the treatment
PS Claim 12: Figure 2; 107pp: English.
CC The full-length TPKS-encoding DNA in plasmid pLOA was
CC designated pTPKS100. Splicing of the introns from the DNA
CC sequence and translation of the 914 nt ORF results in a
CC protein of 3038 AAs (R74171) with a mol. wt. of 269,090
CC daltons. Inspection of the TPKS AA sequence for active
CC site residues and motifs known to be associated with
CC polyketide synthases and fatty acid synthase (FAS) activities
CC resulted in the identification of candidates for expected
CC sites (see FT). Except for the presence of a methyl transferase,
CC not present in FAS, the succession of activities on the
CC TPKS protein is the same as that observed for the rat FAS
CC protein.
SQ Sequence 3038 AA;

```

alignment_scores:

```

Quality: 87.00 Length: 472
Ratio: 0.481 Gaps: 24
Percent Similarity: 38.347 Percent Identity: 19.703

```

alignment_block:

```
US-09-323-427-3 x R74171 ..
```

```
Align seg 1/1 to: R74171 from: 1 to: 3038
```

```

16 ATTGCTTTGTACTACATAATGATGTGTATTGCATTCCGGTGA 65
:::||||| |||::: ||| ||||| |||
1175 ValAlaPhcGlnThrValIleGlyAla.....TyrSerSerPro..... 1187
66 CAATGGTGTGAAGTGCAGCAAAATGTAATG.....GGACCAATT 109
||||: :::::||||| |||||
1188 .....GlyAspArgArgLeuArgCysLeuTyrValProThr 1200

```

```

110 CAATA..... 114
:::
1200 IsValAspArgIleThrLeuValProSerLeuCysLeuAlaThrAlaGlu 1216
115 .....ACAATCAATTTTAATACAGTAATGATTCAGACAGCA 152
::: ||||| |||:::
1217 SerGlyCysGlyIuSValAlaPheAsnThrIleAsnThrTyrAspGly 1233
|||::: |||
1233 ysArgTyrLeuSerGlyAspIleValAlaPheAspAlaGlu..... 1246
|||::: |||
194 ATGATGAAGGTGCACGTCAAGTTCGGCAATTCACCT...CCATTGAT 240
::: ||||| ::|||::: |||||:::
1247 ..GlnThrThrLeuPheGlnValGlnAsnIleThrPheLysProPheSer 1262
241 TCATGCAATGTGGCGGTACAGCATCTGCAATCCACGTGGATTTT... 288
::: |||
1263 ProProAspAlaSerThrAspHisAlaMetPheAlaArgTrpSerTrpG 1279
289 ....GTAACAACAATGTTGTCATTGCTTCATCCATTAATTGTTACCA 334
::: |||
1279 yProLeuThrProAspSerLeuLeuAspAsnProGluTyrTrpAlaThr 1296
335 AAGTTGATCGT.....GCATATGAGTACAAATGCTTTTAC 369
|||::: |||
1296 IacGlnAspGlyGluAlaIleProIleIleGluArgIleValTyrPheTyr 1312
|||::: |||
370 ATGGAAGCTGATAAACAGTGTAGTCAGACAGATTGAGGTATCGAAATCAC 419
::: |||
1313 Ile.....ArgSerPheLeuSerGlnLeuThrLeuGlnGluArgG 1326
420 AACTGCT.....TTTCAACTCAATGTGCCGATGCCAGTATGCGC 460
||| ||| |||||
1326 nGlnAlaAlaPheHisLeuGlnHisLeuGlnIleGluTyrPheGlu..... 1340
511 ATCATTTGGTCACGCCAGTTATGATTAATGACATGCGATTCGAAACCGT 560
||| |||
1351 .....HisLeuTyrTyrAspProGluTyrGlnAs 1360
561 TGATACTTTCGCGCGGTGCATTCCTGTTTGCATGAT..... 603
||||| ::: ||| ||| :::::
1360 nAspThrGluAlaGlnIleGlnHisLeuCysThrAlaAsnSerTyrHisP 1377
603 ..... 603
1377 roHisValArgLeuValGlnArgValGlyGlnHisLeuLeuProThrVal 1393
604 ....GGTAAGCGTGAATCTGGAATTTGAATGCGATGATGCGTCT 650
1394 ArgSerAsnGlyAsnProPheAspLeuAspHisAspGly...LeuLe 1409
651 TGATAAATATTGCTAATAAT.....TGGATAATAC 682
| ::::: |||::: |||:::
1409 uThrGlnPheTyrThrAsnThrLeuSerPheGlyProAlaLeuHisTyrAl 1426
683 CAACAGATTTAATGCGTGGCAAGAACTCAGCTATACAAATATGCGGAT 732
::: |||::: ||| ||||| |||
1426 IaArgGlnLeuValAla...GlnIleAlaHisArgTyrGlnSerMetAsp 1441
733 CGATCAGACGTTTCTATGCAATGCCAGATCAGTATTAACATTAAGAAC 782
1442 IleLeuGlnIle..... 1445
783 AATAGCGAATGTGTTCGACACAATGTTCAAGACCAAGGATTCGAG 832
1446 .....GlyAlaGly. 1448

```



```

833 CTGTTAAACAGGTGGTGGCCGACAGAAAACCTGTCAGCTGGCACTT 882
      |||||
1449 .....ThrGlyGlyAlaThrIleTyrValLeuAlaThrProGlnLeu 1462
882 .....
1463 GlyPheAsnSerTyrThrThrAspIleSerThrGlyPheGluGlu 1479
882 .....
1479 nAlaArgGluGlnPheAlaProPheGluAspArgMetValPheGluProL 1496
883 .....CGTTACTCAGAGAAAAGATCTGCAGACCGAGAAATATC 921
      |||
1496 euAspIleArgSerProAlaGluGlnGlyPheGluPro...HisAla 1511
922 ATTGATGATGACAACTGATATCAACACCCCTTGAATTAAGCATGATATCA 971
      |||||
1512 TyrAspLeuIleIleAlaSerAsnValLeuHisAlaThrProAsp.... 1526
972 AGCTTGCCAGTGTATTTACGTCAC...CGTGCACTTGCAGACATTAATG 1018
      |||
1527 ....LeuGluIleTyrThrMetAlaHisAlaArgSerLeuLeuLysProGly 1542
1019 GACAACTCTAATACCTT.....CGTGCACTACAAATGGAT 1055
      |||||
1542 LysIleMetValIleLeuGluIleThrHisLysGluHisThrArgLeuGly 1558
1036 .....
1559 PheIlePheGlyLeuPheAlaAspTrpTrpAlaGlyValAspAspGlyArg 1575
1056 CTGCATGTCACCATTT 1071
      |||
1575 gCysThrGluProPhe 1580
seq_name: A_Geneseq_36:R06427

seq_documentation_block:
ID R06427 standard; protein; 944 AA.
AC R06427;
DT 14-DEC-1990 (first entry)
DE Protein encoded by sequence which hybridises with DNA encoding
   alpha amidating enzyme.
KW Alpha amidating enzyme; peptidyl amide.
PN EP-382403-A.
PD 16-AUG-1990.
PR 01-FEB-1990; 301034.
PR 06-FEB-1989; US-307366.
PI (UNIG-) UNIGENE LAB INC.
PI Betelsen AH, Mehra NM, Beaudry GA;
DR WPT: 90-248308/33.
DR N-ESDB: Q05636.
PT Prodn. of alpha amidating enzyme - using prokaryotic or
   eukaryotic expression vectors having transcriptional promoter
   Claim 3; page 19; 24pp; English.
CC The corresp. DNA sequence hybridises, under stringent conditions,
   with a DNA sequence (I) which encodes an alpha amidating enzyme
   (AEE) and is foreign to the host cell into which it is trans-
   formed. (I) is connected to a promoter and is contained in an
   expression vector. The AEE encoded is used as a catalyst in the
   conversion of a peptidyl substrate to a corresp. peptidyl amide.
   This can be used for making a protein biologically active, e.g.
   calcitonin or growth hormone releasing factor. See also Q05637.
CC Sequence 944 AA.

```

```

alignment_scores:
  Quality: 86.50      length: 339
  Ratio: 0.588      Gaps: 14
Percent Similarity: 43.363      Percent Identity: 17.994

```

alignment_block:

```

US-09-323-427-3 x R06427 ..
Align seg 1/1 to: R06427 from: 1 to: 944

97 TGTGACCAACTTCATTAACATCAATTTTAATACAGTATGCAATTCGA 146
      ||| |||||
634 CysGlnProThrAspValAlaValGluProSerThr..... 645
147 AGGACATGTTTATGTGAAGGCTTTATGATCAAGAGCGTTGCCGTAATG 196
      ||| |||||
646 .GlyAlaValPheValSerAspGlyTyr..... 654
197 ATGAAGGTGACCTCAAGTTGCCGGAATTCACCTTCATTTGATGATGC 246
655 .....Cys 655
247 AATGTTGCCGTCACAGATCTGCAATCCAGCGTATTTTGTAAACAC 296
      ||| |||||
656 Asn...SerArgIleValGlnPheSerProSerGlyLysPheValThrGlu 671
297 A..... 297
671 nTrpGlyGluGlnSerSerGlySerSerProArgProGlyGlnPheSerV 688
298 .....ACTGTTCATTTGCTTCATTCATTAATTTGTACCAAGTT 339
688 aProHisSerIleAlaLeuValProHisIleuAspGlnLeuCysValAla 704
340 GATCGTGCATAT...CGAGTACAAATGCTTTACATGGAAGCTGATAAAC 386
      |||||
705 AspArgGluAsnGlyArgIleGlnCysPheLysThrAspThrLysGluPh 721
387 AGTTAGTGCACAGATGAGGTATCTGAATCAACACACTGCTTCAAAATC 436
721 eValArgGluIleLysHisAlaSerPheGlyArgAsnValPheAlaIle 738
437 AATTTGCCGATGCCAGTATGCCGTATGAATTTTGATGGTGCACCA 486
738 eTyrIlePro..... 741
487 ACCGGTCACACAGTTCATTTGCTATCATTTGTCAGCCAGTTATGATAA 536
742 .....GlyPheLeuPheAlaValAsnGlyLysProTyrPheGly.. 754
537 ATGACATGCGATTCGAACCTTGATACTTCTCGCGCGTGTGCATTT 586
755 .....AspGlnGluProValGlnGlyPhe..... 762
587 CTTGCTTTGTCAGATGATGTAACGGTATCTGTGAAATTTCTAAATGCT 636
763 .....ValMetAsnPheSerSerGlyGluIleIleAspValPhe..... 775
637 GATGATGTCGCTCTTGATAAATTTGCTAAATTAATTTGGAATATCCAAC 686
776 .....LysProValArgLysHisPheAspMetProH 786
687 AGATTAATATGCTGGCCAGAACCTCAGCTATAC..... 720
786 sAspIleValAlaSerGluAspGlyThrValTyrIleGlyAspAlaHisT 803
721 .....AATATGCGATGATCAGACAGCTTTTCTATTCATATGC 756
803 hTrsAsnThrValTrpLysPheThrLeuThrGluLysMetGluHisArgSer 819
757 .....CAGATCAGTATTAACATTAAGAACAATAAGCAAGATGTGTT.. 798
820 ValLysLysAlaGlyIleGluValGlnGluIleLysAlaGluAlaValA 836
799 .....CGACCACAAATGTTTCAGAACCCAGAGATTCG 829
836 LGIuProLysValGluAsnLysProThrSerSerGlyLeuGln..... 850
830 GAGCTGTTAAACAGGTGTGTCGCGACAGAAAACCTGTCAGCTGGCCAA 879

```

```

851 .....Lys 851
880 CTTGGTTACTCAAGAAAGATCTGCAGACCGAGATATCATTTGATGT 929
852 MetGlnIuLysGlnIuLysLeuSerThrGluProGlySerGlyValSerVa 868
930 ACGAACTGATATCAACACCCCTTGAATTTAGCATGATGATTAATCAAGCTTTC 979
868 I.ValleuIleThrThrLeuLeuValIleProValIleValLeuLeuAla 884
980 CAGTTGATTACGTC 994
885 IleValMetPheIle 889
seq_name: A_Geneseq_36:R10323

```

```

seq_documentation_block:
ID R10323 standard; Protein; 944 AA.
AC R10323;
DR 08-APR-1991 (first entry)
DE Type B alpha-amidating enzyme.
KM Amidation: post-translational modification; AE.
OS Rattus rattus.
PN A09049043-A.
PD 29-NOV-1990.
PF 01-FEB-1990; 049043.
PR 06-FEB-1989; US-307336.
PA (UNIG-) UNIGENE LAB INC.
PI Betelsen AH, Mehra NM, Beaudry GA;
DR WPT:91-022433/04.
DR N-PsDB: Q10278.
PT Expression systems for amidating enzyme - comprises prokaryotic
PT or eurykaryotic hosts contg. a recombinant expression vector contg.
PT the correap. DNA sequence.
PS Disclosure: Page 7; 52pp: English.
CC The sequence was derived from DNA isolated from libraries prepd.
CC from RNA extracted from rat medullary thyroid carcinoma tissues or
CC derived cell lines such as IVI 10028. The cDNAs extracted could
CC be divided into distinct types, the sequence below being encoded by
CC Type B. (Type A is given in R10322). The sequence can be inserted
CC into expression vectors for the prodn. of recombinant alpha-
CC amidating enzyme (AE) useful for post translational modification
CC of other recombinant polypeptides such as calcitonin. The sequence
CC includes a membrane spanning domain which may be undesirable in a
CC recombinant DNA expression system, possibly inactivating the enzyme.
CC A stop codon can be placed upstream of the domain-encoding sequence,
CC pref. between posns. 2340 and 2690 (counting from the CDS). This
CC results in the expression of an 87 rather than a 105 kD prod. The
CC sequences of Type A and Type B are substantially identical with the
CC exception of an Intron region from bases 1178-1492 (CDS) of the
CC Type B cDNA.
SQ Sequence 944 AA:

```

```

alignment_scores:
Quality: 86.50 Length: 339
Ratio: 0.588 Gaps: 14
Percent Similarity: 43.363 Percent Identity: 17.994

```

```

alignment_block:
US-09-323-427-3 x R10323 ..

```

```

Align seg 1/1 to: R10323 from: 1 to: 944
97 TGTGGACCAACCTTCATTAACATTTTAATACAGTAATGCATTGCA 146
||| |||||.....:..|||
634 CysGlnProThrAspValAlaValGluProSerThr..... 645
147 AGGACATGTTTATGGAAGGCTTTATGATCAAGAAGTTGCCGTAATG 196
||| |||:..|||:..|||
646 GlyAlaValAlaPheValSerAspGlyTyr..... 654

```

```

197 ATGAAGGTGGACGTCAAGTTGCCGGAATTTCACTTCATTTGATTCATGC 246
655 .....Cys 655
247 AATGTTCCGGTACACGATCTGTGAATCCAGTGATTTTGTAAACAAC 256
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
656 Asn...SerArgIleValGlnPheSerProSerGlyLysPheValThrG1 671
297 A..... 297
671 nTPGlyGluGluSerSerGlySerSerProArgProGlyGlnPheSerV 688
298 .....ACTGTTGTCATTTGCTTTCATCCATTAATTTGTTACCAAGTT 339
688 alProHisSerLeuAlaIleuValProHisIleuAspGlnIleCysValAla 704
340 GATCGTGCATAT...CGAGTACAATGCTTTTACATGGAACCTGATAAAC 386
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
705 AspArgGluAsnGlyArgGlyIleGlnCysPheLysThrAspThrLysGluPh 721
387 AGTTAGTGCACAGATTGAGGTATCTGAAATCACACAGCTGTTTCAAACTC 436
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
721 eValArgGluIleLysHisAlaSerPheGlyArgAsnValPheAlaIleS 738
437 AAATGTGCCGATGCCAGTATGCCGTTATGAATTTTGGATGCTGACCA 486
: : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
738 eTyrIlePro..... 741
487 ACCGGTCAACCAATTGCATTTGCTATCATTTGTCAGCAGCATTTATCAT 536
742 .....GlyPheLeuPheAlaValAsnGlyLysProTyrPheGly.. 754
537 ATGACATGCGATTTGGAACCGTGTATACTTCTCGCGGTTGTCATTT 586
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
755 .....AspGlnGluProValGlnGlyPhe..... 762
587 CTTGCTTTGCGATGATGTAACCGTGATCTGTGCAAAATCTTAATGCT 636
: : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
763 .....ValMetAsnPheSerSerGlyIleIleLeuAspAlaIlePhe..... 775
637 GATGATGTGCTCTGTATTAATTTGCTTAATTAATTTGGAATATCCAAC 686
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
776 .....LysProValArgLysHisPheAspMetProH1 786
687 AGATTTAATGCTGCGCCACAGACCTCACGTATAC..... 720
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
786 sAspIleValAlaSerLysAspGlyThrValTyrIleGlyAspAlaHisT 803
721 .....AAATAGCGGATGATGATGATGATGATGATGATGATGATGATG 756
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
803 hAsnThrValTyrLysPheThrLeuThrGluLysMetGluHisArgSer 819
757 .....CAGATCAGTATTACCATTAAGAACAATAAGCAATGTGT... 798
: : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
820 ValLysLysAlaGlyIleGluValGlnGluIleLysAlaGluValAlaVa 836
799 .....CGACCACAATGTTGAGAACCAAGGATTCG 829
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
836 GluProLysValGluAsnLysProThrSerSerGluLeuGln..... 850
830 GAGCTGTTAAACAGGTGTGCGCGACAGCAAAACCTGTCAGCTGCGCAA 879
851 .....Lys 851
880 CTTGTTTACTCAAGAAAGATCTGCAGACCGAGATATCATTTGATGT 929
852 MetGlnIuLysGlnIuLysLeuSerThrGluProGlySerGlyValSerVa 868
930 ACGAACTGATATCAACACCCCTTGAATTTAGCATGATGATTAATCAAGCTTTC 979
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
868 I.ValleuIleThrThrLeuLeuValIleProValIleValLeuLeuAla 884
980 CAGTTGATTACGTC 994

```

885 ILeValMetPheIle 889

seq_name: A_Geneseq_36:W13009

seq_documentation_block:

ID W13009 standard; protein: 560 AA.

AC W13009;

DT 21-NOV-1997 (first entry)

DE Segment of desmosomal cadherin, desmoglein Dsg2.

KW Desmosomal cadherin; desmoglein; Dsg2; cell surface; epithelial;

KW Carcinoma; desmosome; antibody; epitope; diagnosis; detection;

KW micrometastasis; separation; enrichment; targeted delivery;

KW metastatic.

OS Homo sapiens.

PN DE19531033-A1.

PD 27-FEB-1997.

PF 23-AUG-1995; 031033.

PA (PROG-) PROGEN BIOTECHNIK GMBH.

PI Franke WW, Schaefer S;

DR WPI: 97-146518/14.

PT Antibody reactive with part of desmosomal cadherin - exposed on surface of epithelial or carcinoma cells, not bound to desmosomes, useful for diagnosis and treatment of carcinoma micrometastases

PS Claim 7; Page 5; 8pp; German.

CC The present sequence is a segment of the desmosomal cadherin (DC), desmoglein Dsg2, which is exposed on the surface of epithelial or

CC carcinoma cells and not bound to desmosomes. An antibody (Ab) directed against epitopes of the present sequence can be used to

CC diagnose, i.e. to detect carcinoma cells, especially

CC micrometastases, not bound to desmosomes, to separate, enrich or

CC detect living or fixed carcinoma cells by cell sorting methods and

CC as a therapeutic to deliver agents, e.g. other Ab or toxins, to

CC target cells. The Ab provides rapid and reliable detection of

CC metastatic carcinoma, and detects parts of DC that are not

CC accessible in desmosome bound cells, as in normal tissue or

CC carcinomas.

SQ Sequence 560 AA:

alignment_scores:

Quality: 85.50 Length: 203

Ratio: 0.822 Gaps: 12

Percent Similarity: 51.232 Percent Identity: 25.616

alignment_block:

US-09-323-427-3 x W13009 ..

Align seg 1/1 to: W13009 from: 1 to: 560

```

43 TTGCTTATTCGATTCGCGTTGACATGCTGCAAGGTGAGCCAGCAAT 92
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
299 LeuAspPheSerValIleValAlaAsnLysAlaAlaPheHispsrTl 315
   93 T.....GAATGTGACCACTTCATACATACATATTTATACAGTA 136
   |       |||:|||||:|||||:|||||:|||||:|||||:
315 ArgSerLysTyrLysProThrProIleProIleLysValLysVal 332
   137 ATGCATTCCAGCA.....CATGTTTATGTG 162
   ||:|||||:|||||:|||||:|||||:|||||:
332 snValLysGluGlyIleHisPheLysSerValIleSerIleTyrVal 348
   163 AAAGTCTTTATTCATCAAGAAGTTCGCTAATGATGAGTGCAGCTCA 212
   |||:|||||:|||||:|||||:|||||:|||||:
349 SerGluSerMetAsp.....ArgSerSerLysGly.....GI 359
   213 AGTTGCCGGAATTTCACTTCATTGATTCATGCAT.....G 250
   ||:|||||:|||||:|||||:|||||:|||||:
359 nIleIleGlyAsnPheGlnAlaPheAspGluAspThrGlyLeuProIaH 376
   251 TTGCGCTACACGATCTGTGATTCACGCTGATTTT..... 288
   |||:|||||:|||||:|||||:|||||:

```

seq_name: A_Geneseq_36:R20026

seq_documentation_block:

ID R20026 standard; protein: 973 AA.

DT 08-APR-1992 (first entry)

DE C-terminal amidation enzyme.

KW Horse.

OS Equus caballus.

FT key Location/Qualifiers

FT region 829..895

FT J03262484-A. /note="deleted in R20027 and R20028"

PD 22-NOV-1991.

PF 14-MAR-1990: 063306.

PR 14-MAR-1990: JP-063306.

PA (SHTS) SHISEIDO KK.

DR WPI: 92-012701/02.

DR N-PSDB: Q20198.

PT cDNA sequence - coding for peptide C terminal amidation enzyme

PS Claim 1; Fig 1; 18pp; Japanese.

CC The sequence was deduced from the DNA sequence determined from a

CC clone isolated from a library prepd. from mRNA extracted from

CC horse atrium tissue. The sequence is one of four similar ones

CC provided which all have the same sequence up to residue 809 at

CC which point they diverge, having different C-termini. The

CC different termini are created by deletions in the last portion of

CC DNA (see feature table).

CC See also R20025-28.

SQ Sequence 973 AA:

alignment_scores:

Quality: 85.00 Length: 332

Ratio: 0.582 Gaps: 14

Percent Similarity: 43.976 Percent Identity: 18.072

alignment_block:

US-09-323-427-3 x R20026 ..

Align seg 1/1 to: R20026 from: 1 to: 973

```
97  TGTGACCACTTCATTAACATCATATTTAAACAGTAATGCATTGCA 146
    ||| |||||:~::~:~::|||
630  CysGlnProThrAspValAlaValAspProAsnThr..... 641
147  AGGACATGTTTATGTGAAGGCTTTATGATCAAGAAGTTGCCGTATG 196
    ||| ~::~:~::|||
642  .GlyThrIlePheValSerAspGlyTyr..... 650
197  ATGAGGTGAGCTCAAGTTCGCGAATTTCACTTCATTTGATTATGC 246
    ||| ~::~:~::Cys 651
651  ..... 651
247  AATGTGCGGCTACAGCATCTGAAATCAGCTGATATTTGTATCAAC 296
    ||| ~::~:~::||| |||||
652  Asn...SerArgIleValGlnPheSerProThrGlyArgPheIleThrGI 667
297  A..... 297
667  nTPGlyGlnGluSerSerGluSerAsnProLysProGlyGlnPheArgV 684
298  .....ACTGTTGTCATTTGTTTCATTCATTAATTTGTACCAAGTT 339
    ||| ~::~:~::||| ~::~:~::
684  aProHisSerIleuAlaIleuValProHisIleuGlyIleuCysValAla 700
340  GATCGTGCATAT...CGAGTACAATGCTTTTACATGGAAGCTGATAAAC 386
    ||||| ~::~:~::||| ~::~:~::
701  AspArgGluAsnGlyArgIleGlnCysPheIleThrAspThrLysGluPh 717
387  AGTTAGTGCACAGATGAGTATCTGAAATCACAACGCTTTTCAACATC 436
    ||| ~::~:~::||| ~::~:~::
717  eValArgGluIleLysHisAlaSerPheGlyArgAsnValPheAlaIleS 734
437  AAATTCGCCGATGCCAGTATGCCGTTATGAATTTGGATGCGGACCA 486
    ~::~:~::|||
734  eTyrIlePro..... 737
487  ACCGGTCACCACTTCATTTGCTATCATGTCACGCCAGTTATCATTA 536
    ~::~:~::||| ~::~:~::
738  .....GlyLeuLeuPheAlaValAsnGlyLysProTyrPheGly.. 750
537  ATGGACATGGGATTTGGAACCGTTGATCTGTCGGCGGTGTGCAT 586
    ||| ~::~:~::|||
751  .....AspGlnLysProValGlnGlyPhe..... 758
587  CCTGCTTTGTCATGATGATGACGATGATGATGTAATTCATAATGCT 636
    ~::~:~::||| ~::~:~::
759  ....ValMetAsnPheSerSerGlyLysIleIleAspValPhe..... 771
637  GATGATGTGCTCTTGATTAATATTGCTTAATAATTGCAATATCCAC 686
    ||| ~::~:~::|||
772  .....LysProValArgLysHisPheAspMetProHI 782
687  AGATTTAATGGCTGCCAAGACGTCACGTATAC..... 720
    ||||| ~::~:~::|||
782  sAspIleThrAlaSerGluAspGlyThrValIleValGlyAspAlaHisT 799
721  .....AAATATGCGATGATCAGACGCTTTCTATCAATGC 756
    ||||| ~::~:~::|||
799  hrasnThrValITrPlysPheThrSerThrGluArgValGluHisArg... 814
757  CAGATCAGTATTACCATTAAGAACAATAGCAATGTTGCAACCA 806
    ~::~:~::||| ~::~:~::
815  .....SerValLysLysAlaGlyIleGluValGlnGluIleL 827
807  ATGTTGAGAACCAAGAGATTGCGAGCTGTTAAACAGGTGTCGCCAG 856
    ~::~:~::|||
827  sGluSerGlu.....AlaValValGluThrLysMetGluA 839
857  CAAACCTGCTGCAGCTGCCAATTCGTTTACTCAGAAAGATCT... 903
    ||||| ~::~:~::||| ~::~:~::
839  snLysProAlaSerSerGluLeuGlnLysMetGlnLysGlnLysLeu 855
904  ...GCAGAACCGGAGAAATATCATGATGTAAGACGATATCAACACCT 950
    ||||| ~::~:~::||| ||||| ~::~:~::|||
856  IleLysGluProGlySerGlyValProVal.ValLeuIleThrThrLeuL 872
951  TGAATTTAGCGATGATATATCAAGCTTTGCCAGTTGATTTACGTC 994
    || ~::~:~::||| ~::~:~::
872  euValIleProValValValLeuValAlaIleAlaIlePheIle 886
```

11:11:15

THIS PAGE BLANK (USPTO)

OM of: US-09-323-427-3 to: Issued_Patents_AA.* out_format: pfs

Date: Apr 16, 2000 4:46 AM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frimex.n2p.model -DEV=xip
-O7/cgn2_1/USPTO.spool/US09323427/runat.14042000.170513.19887/app_query.fasta.1
-DB=Issued_Patents_AA -QFM=fastan -SUFFIX=backtrans.ra1
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPECI=0.000
-LOOPEXT=0.000 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blonsum62 -TRANS=human40.cdi -LIST=45 -DOCCALIGN=200
-THR SCORE=pcr -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=next
-MILEN=0 -MAXLEN=1000000 -USER=US09323427 -NCPU=6 -ICPU=3
-NO_XLPRY -WAIT -THREADS=1

Search information block:

Query: US-09-323-427-3
Query length: 1161
Database: Issued_Patents_AA.*
Database sequences: 13390
Database length: 13297546
Search time (sec): 99.320000

Score list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation	
/cgn2_6/ptodata/1/1aa/5A.COMB.pep:US-07-707-367-2 +		100.00	163.44	0.0075	935	1
/cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-070-301-16 +		96.50	155.51	0.0195	989	1
/cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-477-451-8 +		96.50	141.07	0.0385	3200	1
/cgn2_6/ptodata/1/1aa/5A.COMB.pep:US-07-906-349A-6 +		93.00	150.86	0.0438	801	1
/cgn2_6/ptodata/1/1aa/5A.COMB.pep:US-08-415-751-35 +		91.50	157.52	0.0412	362	1
/cgn2_6/ptodata/1/1aa/5A.COMB.pep:US-08-325-071-67 +		90.00	147.22	0.0861	650	1
/cgn2_6/ptodata/1/1aa/5A.COMB.pep:US-08-325-071-61 +		89.00	147.22	0.1018	549	1
/cgn2_6/ptodata/1/1aa/5A.COMB.pep:US-08-325-071-63 +		89.00	145.15	0.1123	650	1
/cgn2_6/ptodata/1/1aa/5A.COMB.pep:US-08-450-332-2 +		87.00	122.05	0.4648	3038	1
/cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-637-640-2 +		87.00	122.05	0.4648	3038	1
/cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-540-804-12 +		85.50	130.11	0.4099	1226	1
/cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-218-265-12 +		85.50	130.11	0.4099	1226	1
/cgn2_6/ptodata/1/1aa/5A.COMB.pep:US-08-405-604A-6 +		85.00	119.32	0.7398	2710	1
/cgn2_6/ptodata/1/1aa/5A.COMB.pep:US-08-450-496A-6 +		85.00	119.32	0.7398	2710	1
/cgn2_6/ptodata/1/1aa/5A.COMB.pep:US-08-325-071-65 +		84.00	135.39	0.4118	620	1
/cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-070-301-14 +		83.50	128.61	0.6157	989	1
/cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-682-517-15 +		83.00	124.98	0.7944	1222	1
/cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-682-517-9 +		83.00	124.43	0.8056	1252	1
/cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-162-081B-36 +		82.00	124.43	0.9645	1080	1
/cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-780-872-36 +		82.00	124.43	0.9645	1080	1
/cgn2_6/ptodata/1/1aa/5A.COMB.pep:US-08-348-006B-5 +		81.50	116.37	1.53	1911	1
/cgn2_6/ptodata/1/1aa/5A.COMB.pep:US-08-800-825A-5 +		81.50	116.37	1.53	1911	1
/cgn2_6/ptodata/1/1aa/PCUS9.COMB.pcr:US94-10166-5 +		81.50	116.37	1.53	1911	1
/cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-477-451-3 +		81.50	110.13	2.05	3174	1
/cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-742-440A-3 +		81.00	134.36	0.7161	407	1
/cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-162-081B-37 +		81.00	122.48	1.25	1069	1
/cgn2_6/ptodata/1/1aa/5A.COMB.pep:US-08-415-751-6 +		80.50	134.76	0.7643	362	1
/cgn2_6/ptodata/1/1aa/5A.COMB.pep:US-08-785-076-3 +		80.50	132.65	0.8441	430	1
/cgn2_6/ptodata/1/1aa/5A.COMB.pep:US-08-325-071-59 +		80.50	132.53	0.8464	432	1
/cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-484-993B-16 +		79.00	126.08	1.48	570	1
/cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-484-158B-16 +		79.00	126.08	1.48	570	1
/cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-484-569A-16 +		79.00	126.08	1.48	570	1
/cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-480-150A-16 +		79.00	126.08	1.48	570	1
/cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-450-731-16 +		79.00	126.08	1.48	570	1
/cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-070-301-17 +		79.00	119.30	2.03	989	1
/cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-484-575A-20 +		78.50	128.39	1.44	434	1
/cgn2_6/ptodata/1/1aa/5B.COMB.pcr:US-08-477-459A-20 +		78.50	128.39	1.44	434	1
/cgn2_6/ptodata/1/1aa/PCUS9.COMB.pcr:US94-01826A-20 +		78.50	128.39	1.44	434	1
/cgn2_6/ptodata/1/1aa/PCUS9.COMB.pcr:US94-02252A-20 +		78.50	128.39	1.44	434	1

/cgn2_6/ptodata/1/1aa/PCUS9.COMB.pcr:US96-03916-11 + 78.50 128.39 1.44
/cgn2_6/ptodata/1/1aa/PCUS9.COMB.pcr:US96-03916-67 + 78.50 128.39 1.44
/cgn2_6/ptodata/1/1aa/5B.COMB.pcr:US-08-819-288A-3 + 78.50 114.96 2.71
/cgn2_6/ptodata/1/1aa/5A.COMB.pcr:US-08-261-822A-3 + 78.50 114.71 2.74
1321

seq_documentation_block:

Sequence 2, Application US/07707367

Patent No. 5196316

GENERAL INFORMATION:

APPLICANT: Iwasaki, Yasuno

APPLICANT: Shimoi, Hiroko

APPLICANT: Suzuki, Kenji

APPLICANT: Gishalpa, Oreste

APPLICANT: Nishikawa, Yoshiki

APPLICANT: Kawahara, Takashi

APPLICANT: Kangawa, Kenji

TITLE OF INVENTION: No. 5196316el Enzyme and DNA Coding Therefor

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESS: CIBA-GEIGY Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: New York

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

SOFTWARE: PC-DOS/MS-DOS

OPERATING SYSTEM: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/707,367

FILING DATE: 19910530

CLASSIFICATION: 530

Prior APPLICATION DATA:

APPLICATION NUMBER: JP 141678/90

FILING DATE: 01-JUN-1990

Prior APPLICATION DATA:

APPLICATION NUMBER: JP 210535/90

FILING DATE: 10-AUG-1990

Prior APPLICATION DATA:

APPLICATION NUMBER: JP 329911/90

FILING DATE: 30-NOV-1990

ATTORNEY/AGENT INFORMATION:

NAME: Williamizar, Joann

REGISTRATION NUMBER: 30,598

REFERENCE/DOCKET NUMBER: 4-18110/A/CGJ 44

TELECOMMUNICATION INFORMATION:

TELEPHONE: (914)785-7120

TELEFAX: (919)541-8689

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 935 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-707-367-2

alignment_scores:

Quality: 100.00

Ratio: 0.541

Percent Similarity: 47.315

Percent Identity: 19.437

alignment_block:

US-09-323-427-3 x US-07-707-367-2 ..

Align seg 1/1 to: US-07-707-367-2 from: 1 to: 935

97 TGTGGACCACTTCATACATCAATTTTAATACGATATCATTCGA 146

|||||..... ||| |||

```

526 CysGlnProThrAspValAlaValAspProIleThrGlnAsnPhenPheVal 542
147 AGGACAGTTTATGTGAAGGTCTTTATGATCAAGAAGTTGCCGTATG 196
542 LAlaAspGlyTyr..... 546
197 ATGAAGGTGAGCTCAAGTTGCCGAATTCACATTCATTGATTCATGCC 246
547 ..... 547
247 AATGTTCCGGTACAGATCTCGAATCCAGCGTGATTTTGTGA..... 291
548 Asn...SerArgIleMetGlnPheSerProAsnGlyMetPheIleMetG1 563
291 ..... 291
563 nTPrGlyGlnGluThrSerSerAsnValProArgProGlyGlnPheArg1 580
292 .....ACACAGACTGTGTGCAATTTGTTTCATTCATTATTTGTACC 333
580 IeProHisSerIeuThrMetValProAspGlnGlnLeuGlnValAla 596
334 AAGGTGATGTCGATATCGAGTACAAATGCTTTACATGGAAGCTGTAA 383
597 AspArgGlnAsnGly...ArgIleGlnCysPheHisAlaGluThrGlnGlyAs 612
384 AACAGTTAGTCACAGATGTGAGTATCGTAATC...ACACAGCTTTTC 430
612 nPheValGly...GlnIleGlyHisGlnGlnPheGlyArgGluValAlaPhea 628
431 AAACCTCAATTTGTCGGATGCCAGTATGCCGTTATGAAATTTTGATGAT 480
628 lValIleSerTyrAlaPro.....Gly 634
481 GGAACAACGGGTCAACAGTTCAATTTGTCATGTCATGTCAGACGATTTA 530
635 Gly.....ValLeuTyrAlaValAsnGlnGlyAsnProTyrTy 646
531 TCATAAATGACATGCGATTTGAAACCGTTGATCTTCTGCGCGGTTG 580
646 r.....GlyTyrSerAlaProV 652
581 TCCATTCTCTCTTTGTCGATGATGTAACCGGTATCTGGAATTTCTA 630
652 aIcGlnGlyPheMetLeuAsnPhSerAsnGlyAsp..... 663
631 AATGCTGATGATGCTGCTGTGAATAATATTG.....CTAATATATT 674
664 .....IleLeuAsnProThrPheIleProAlaArgGlyAsnPh 675
675 GGAATATCCACAGATTTAATGCTGTCGCAAGAAGCTCAGGTATACAAAT 724
675 eAspMetProHisAspIleAlaAlaAspAspGlyThrValTyrValG 692
725 ATGCGGATGATCAGACTTTTATATCATGCCAGATCAGT..... 765
692 lAspAlaHisAlaAsnAlaValTyrPheSerProSerGlyAlaGlu 708
766 ..ATTACGATTAAGAACCAAT.....AGCGAATGCTGTGAC 803
709 HisArgSerValGlyAsnAlaGlyIleGluValGlnGluIleThrGluTh 725
804 ACAATGTTGACAGACAGATTCGAGCTGTTAAACAGGTGTCGCCG 853
725 rGluIlePheGlnThrHisIleArgSerArgProGlySTHisAsnGlySerV 742
854 CAGCAAAACCTGCTGCAAGTCCCAACTTCGTTTACTCAAGAAAGATCT 903
742 aGlnGlySerGlnThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 756
904 GCA.....GAACGGAGAAATATCATTT...GATGTAGAAC 935
757 AlaGlyValSerThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 773

```

```

936 TGATATCAACACCTTGAA.....ATTAGCATGATATCAAGCTT 976
773 aGlyValProThrGlnGlnGlnGlnAsnValValGlnGlnGlnSerAlaG 790
977 TGCAGTTGATTTAGCTCACCGTGCACCTTCTGCACATAATAGCAACCT 1026
790 lValSerThrGlnGlnGlnGlnGlnSerValValGlnGlnGlnSerAlaGly 806
1027 GTAATACTTCTGCAGTACAAATGCAATCGATCGATCCATTTGGCTT 1076
807 ValSerThrGlnGlnGlnGlnGlnSerValValGlnGlnGlnSerAlaGly 823
1077 CTCAAATGTTATGGTTTAAAGCATTCATGATT...GCTGCCGTCATTA 1123
823 lSerPheValIleuIleIleThrIleuIleIleProIleAlaValIleu 840
1124 TTACCATTTGTTAAATTTGCT 1146
840 lValIleAlaIlePheIleArg 847

seq_name: /cgn2.6/plodata/1/1aa/5B_COMB.pep:US-08-070-301-16
seq_documentation_block:
: Sequence 16, Application US/08070301
: Patient No. 5871995
: GENERAL INFORMATION:
: APPLICANT: ITDA, Toshio
: APPLICANT: KAMINUMA, Yoshihiko
: APPLICANT: FUSE, Yuka
: APPLICANT: TAJIMA, Masahiro
: APPLICANT: YANAGI, Mitsuo
: APPLICANT: OKAMOTO, Hiroshi
: APPLICANT: KISHIMOTO, Jiro
: APPLICANT: IFUKU, Ochi
: APPLICANT: KATO, Ichiro
: TITLE OF INVENTION: ENZYME PARTICIPATING IN C-TERMINAL
: NUMBER OF INVENTION: AMIDATION, AND METHOD OF PREPARING SAME AND USE THEREOF
: CORRESPONDENCE ADDRESS:
: ADDRESS: Wegner, Cantor, Mueller & Player, P.C.
: STREET: 1233 20th Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20036-8218
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/070,301
: FILING DATE: 24-MAY-1991
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 1-209687
: FILING DATE: 15-AUG-1989
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 1-181933
: FILING DATE: 31-OCT-1989
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 2-76331
: FILING DATE: 26-MAR-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 2-106412
: FILING DATE: 24-APR-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 2-205475
: FILING DATE: 02-AUG-1990
: ATTORNEY/AGENT INFORMATION:
: NAME: Player, William E.
: REGISTRATION NUMBER: 31,409

```



```

REFERENCE/DOCKET NUMBER: P-450-22830
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-040
TELEFAX: (202) 835-0605
TELEX: 440706
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 989 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Frog
US-08-070-301-16

alignment_scores:
Quality: 96.50      Length: 232
Ratio: 0.965       Gaps: 12
Percent Similarity: 43.103   Percent Identity: 21.552

alignment_block:
US-09-323-427-3 x US-08-070-301-16 ..

Align seg 1/1 to: US-08-070-301-16 from: 1 to: 989

97 TGGGACCACTTCATATACATCAATTTAATACAGTAATGATTCGA 146
||| |||||.....:||||| ||| |||
642 CysGlnProThrAspValAlaValAspProIleThrGlyAsnPhenIle 658
147 AGGACATGTTATGTGAAGGCTTTATGATCAAGAAGTTGCCGTATG 156
::: |||
658 LAlaAspGlyTyr.....:..... 662
197 ATGAGGTGACGTCGAAGTCCGGAAATTCCTCATTTGATTCATGTC 246
|||
663 .....:.....Cys 663
247 AATGTCGGGTACACAGATCTGATCCACAGCGTATTTTGTGA.... 291
||| .....:.....:|||||
664 Asn.....SerArgIleMetGlnPheSerProAsnIleMetIleMetG1 679
291 .....:..... 291
679 nTPGlyGluGluThrSerSerAsnLeuProArgProGlyGlnPheArg1 696
292 .....ACACACTGTTGTGTCATTTGCTTCATTCATTAATTTGTTACC 333
::: |||||.....: ||| |||
696 IeProHisSerLeuThrMetIleSerAspGlnGlyGlnLeuGlyValAla 712
334 AAGTTGATGTCGATATGATGATGATGCTTTTACATGAGCGTGATAA 383
:.....:.....:.....:
713 AspArgGlnAsnGly...ArgIleGlnCysPheHisAlaGlySthnGlyG1 728
384 AACAGTTAGTCACAGATGAGTATGTAATC...ACAACGCTCTTTC 430
: |||||.....: |||||
728 uphLeuValGly...GlnIleGlySthnGlnIleGlyValGlyValAlaPhe 744
431 AAATCAATGTCGCCGATGCCAGTATGCCGTTATGATAATTTGGATGCT 480
:.....:.....: |||
744 lAlaSerTyrAlaPro.....:.....Gly 750
481 GGACCAACCGGTCAACCATTCATTTGCTATCATTTGGTCAGCGCATTTA 530
||| .....:.....: |||||
751 Gly.....ValLeuTyrAlaValAlaGlnGlySthnGlySthnGly 762
531 TCATTAATGACATGCGATTCGTGAACCGTTGATCTTCTGCGCGGCTTG 580
: ||||| .....: |||||
762 rGly.....AspSerThrProValGlnGlyPhe..... 771
581 TCATTTCTCTTGTGTCATGATGATGATGATGATGATGATGATGATGAT 630
:.....:.....: |||||

```

```

772 .....MetIleAsnPhSerAsnGlyAsp..... 779
631 AATGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 674
780 .....IleLeuAspThrPheIleProAlaArgIlySthn 791
675 GGAATATCCAAGATTTAATGCTGCGCAAGAAGCTCAGCTATATC 720
: ||| ||| .....: |||||
791 eGlnMetProHisAspIleAlaValAlaGlyAspAspGlyThrValTyr 806

seq_name: /cgn2_6/plodata/1/1aa/5B_COMB.pep:US-08-477-451-8

seq_documentation_block:
; Sequence 8, Application US/08477451
; Patent No. 5928865
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori CagI Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,451
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0335.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-655-3542
; TELEFAX: 510-601-2708
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3200 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-477-451-8

alignment_scores:
Quality: 96.50      Length: 369
Ratio: 0.603       Gaps: 20
Percent Similarity: 43.360   Percent Identity: 22.764

alignment_block:
US-09-323-427-3/rev x US-08-477-451-8 ..

Align seg 1/1 to: US-08-477-451-8 from: 1 to: 3200

1050 ATTTGTACTGCAGCAAGTAT.....TACAGTTGTTCATTTATGTT 1010
|||.....:.....: ||||| |||
745 lIleValPheCysArgIlePheGlnLeuValTyrArgLeuLeuIleAla 761
1009 GCAGAGTCGACGCGTGCAGTAAATCACTGCAAGCTTGATTAATCATCG 960
::: ||| .....: ||||| |||
761 slySthnCys.....PheSerPheTyrArgLeuIleGlnIleVal 774
959 CTAATTTCAAGGCTGTGATATC.....AGTTGTCATCATCAATGATATT 916
: ||||| .....: |||||
774 alaSnPheLeuSerGlnPheLeuIleProIlySthnGlySthnGly 790

```


604 604
 966 TATCAAGCTTGCAGTTGATTACGTCACCGCTGCATTCGACACATA 1015
 ||||| |||||
 605ThrglyThrCysThrThrThrThr 613
 1016 ATGCACACCTGTAAATCTGCTGCAGTACA 1046
 ||||| |||||
 613 rThrThrThrCysAlaThrCysCysThrThr 623

seq_name: /cgn2_6/plodata/1/iaa/5A_COMB.pep:US-08-415-751-35

seq_documentation_block:

: Sequence 35, Application US/08415751
 : Patent No. 5643772

: GENERAL INFORMATION:

: APPLICANT: PETERSEN, CAROLYN

: APPLICANT: LEECH, JAMES

: APPLICANT: NELSON, RICHARD, C.

: APPLICANT: GUT, JIRI

: TITLE OF INVENTION: POLYPEPTIDES BINDING ANTI-

: TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA

: TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID

: TITLE OF INVENTION: VECTOR AND TRANSFORMED HOST AND

: TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND

: NUMBER OF SEQUENCES: 50

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: PHILLIPS, MOORE, LEMPJO & FINLEY

: STREET: 385 Sherman Avenue, Suite 6

: CITY: Palo Alto

: STATE: California

: COUNTRY: United States of America

: ZIP: 94306-1840

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Kb storage

: COMPUTER: PC

: OPERATING SYSTEM: DOS

: SOFTWARE: Wordperfect 5.1

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/415,751

: FILING DATE: 03-APR-1995

: CLASSIFICATION: 435

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: 08/071,880

: FILING DATE: June 1, 1993

: APPLICATION NUMBER: 07/891,301

: FILING DATE: May 29, 1992

: ATTORNEY/AGENT INFORMATION:

: NAME: Hana Dolezalova

: REGISTRATION NUMBER: 30,518

: REFERENCE/DOCKET NUMBER: 480.19-2 (HHD)

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (415) 324-1677

: TELEFAX: (415) 324-1678

: INFORMATION FOR SEQ ID NO: 35:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 352 amino acids

: TYPE: amino acid

: STRANDEDNESS: single

: TOPOLOGY: linear

: MOLECULE TYPE: peptide

: ORIGINAL SOURCE:

: ORGANISM: Cryptosporidium parvum

: FEATURE:

: NAME/KEY: Positions coded by nonsense codons are

: NAME/KEY: identified as Xaa.

: US-08-415-751-35

alignment_scores:

Quality: 91.50

Ratio: 1.220

Length: 151

Gaps: 9

Percent Similarity: 49.669 Percent Identity: 25.828

alignment_block:

US-09-323-427-3 x US-08-415-751-35 ..

Align seg 1/1 to: US-08-415-751-35 from: 1 to: 362

761 TCAGTATTACCATTAAGAACCAATAGGAATGTTGCACAC...AA 807
 || || |||||
 20 SerAsnLeuLeuLeuGlnHisSerArgGlnHisPheGlnArgLeuG1 36
 808 TGTTCAGAACCAACAGAGATTGCGAGCTGTTAAACAGGTGTCGCCAC 857
 ||||| |||||
 36 yMetArgAsn***ProHisArgGlnLeuLeuHisGlnLeuValValLysG 53
 858 AAACCTGCTGCAGCTGCCAC.....TTG 883
 |||||
 53 LThrIle**HisLeuLysAsnArgCysLysTyrGlnSerThrHis 69
 884 GTTACTCAAGAAAGAT.....CTGCAGAACCGAGAGATTCATTGAT 927
 |||||
 70 LeuGlnSerGlnIleAspPheLysLeuGlnAsnArgLeu..... 82
 928 GTAAGACTGATATCAACACACCTTGAATTAGGATGATATCAAGCTT 977
 |||||
 83LeuValGlu****LeuGlnLeuArgLeuValValLeuLeu* 97
 978 GCCAGTTGATTACGTCACCGTGCACCTTGCACACATAATGACACA.... 1022
 || || |||||
 97 ****LeuGlnPhe.AspleuGlyValLeu*****Tyr**ArgLe 113
 1023 .ACCTGTAACTGCTGCTGCAGTACAAATAGGAAT..... 1055
 |||||
 113 uCysCysCysGlyCysCysSerLysIleIleIlePaspAsnCysCysSerL 130
 1056CTGCAGTGCACCATTTGGCTTCTC 1079
 |||||
 130 euPheIlePheCysCysCysCysCysLeuAsnIleIleIleIleIlePhe 145
 1080 AATGTTATGCGTTTAAGCATTTGATTTGCTGCGCATTAATATACAA 1129
 |||||
 146TyrGly.....CysCysCysCysCys**LeuIlePhe 157
 1130 T 1130
 157 n 157

seq_name: /cgn2_6/plodata/1/iaa/5A_COMB.pep:US-08-325-071-67

seq_documentation_block:

: Sequence 67, Application US/08325071

: Patent No. 5587311

: GENERAL INFORMATION:

: APPLICANT: COBON, Stewart Gary

: APPLICANT: MOORE, Joanna Terry

: APPLICANT: JOHNSON, Law Anthony York

: APPLICANT: WILLADSEN, Peter

: APPLICANT: KEMP, David Harold

: APPLICANT: SRISKANTHA, Alagacone

: APPLICANT: RIDING, George Alfred

: APPLICANT: RAND, Keith No. 5587311man

: TITLE OF INVENTION: DNA Encoding A Cell Membrane

: TITLE OF INVENTION: glycoprotein of A Tick Gut

: NUMBER OF SEQUENCES: 71

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Foley & Lardner

: STREET: 3000 K Street, N.W.

: CITY: Washington, D.C.

: COUNTRY: USA

: ZIP: 20007-5109

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/325,071
FILING DATE: 14-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,109
FILING DATE: 17-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/926,368
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/242,196
FILING DATE: 06-JUL-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU87/00401
FILING DATE: 27-NOV-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P14912
FILING DATE: 16-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P12570
FILING DATE: 19-JUN-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P9196
FILING DATE: 27-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 60042/111 BIAU
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 650 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-325-071-67

alignment_scores:
Quality: 90.00 Length: 359
Ratio: 0.612 Gaps: 22
Percent Similarity: 40.947 Percent Identity: 21.727

alignment_block:
US-09-323-427-3 x US-08-325-071-67 ..
Align seg 1/1 to: US-08-325-071-67 from: 1 to: 650

138 TGCATTCGAAGCAGCATCT.....TTATG 160
|||||
147 CysValProThrThrCysLeuArgProAspLeuThrCysLysAspLeu 163
161 TGAAGGTTTATGATCAAGAGTTGCCGTAATGATGAAGTGACG 210
|||||
163 sglutylsasnleuleuglnargasp.....Sera 173
211 CAAGTTCGCGAATTCACATTCATTCATCAGATGCGCGTAC 260
|||||
173 rgcyscysglnlytrpansThr.....AlasncysSeraIaIa 186
261 ACGATCTCGAATCCACGCGTATTTTGTAAAC..... 293
187 ProProIaAspSer.....TyrCysSerProGlySerProLysgl 200
294AACACTGTGTCAAT.....TTGC 312
200 yProAspGlyGlnCysLysAsnIaCysArgThrLysGlnIaGlyPhe 217

313 TTTCAATCATTTATTGTACCAAGTTGATCGCATATCGATACAAATG 362
|||
217 AlCysLysHisGlyCys.ArgSerThrAspLysAlaTylGluCysThry 233
363 C.....TTTACATGAGACGTGATAAACAAGTATGACGACATATG 403
233 sProSerGlySerThrValaIaGlnAspGlyIleThrCysLysSerIle 250
404 ACGTATCTGCAAAATCACAACTGCTTTTCAAACTCAATTGTC...CCGATG 450
|||
250 eTyrThrValSerCysThrValaGlnLysGlnThrCysArgProThr 266
451 CCAGTATGCCGTTATGAA.....ATTTGAGNGTGAGCAACCG 491
267 GluAspCysArgValGlnLysGlyThrValaLeuGlyCysProTrpAs 283
492 TCACACAGTTCAATTTGCTATCATGTCAGCCAGTTTATCAAAATGGA 541
|||
283 ngIn.....HisLeuValGlyAsp.....T 290
542 CATCGATTCGAAACCGTTGATACCTTTCGCGCGTTGTCATTCCTGC 591
|||||
290 hCysIleSerAspCysValAspLysLysCys.....HisGlnGlu 303
592 TTTGTGATGATGCT..... 606
|||
304 PheMetAspCysGlyValTyrMetAsnArgLysSerCysIleCysPro 320
607AACGGTATACGTGGAAATTTCAATTCGATGAGAT 643
320 pLysSerArgLysProGlyProAsnVal.....AsnIleAsnGlu 334
644 GTGCTCTGATAAATATTGCTAAAT.....AATTGGAA 678
|||
334 yLeuLeuAsnGlnTyrTyrThrValSerPheThrProAsnIleSer 350
679 TATCAACAGATTTAATGGCTGCCAAGACCTACAGTATCAAAATATGC 728
|||
351 PheAspSerAsp.....HisCysLysArgTyrGln 360
729 GATGATCATCAGCTTTTCTATCATGCGCAGATCATGATTAACATTAAG 778
|||||
360 uAspArg.....ValLeuGlnAlaIleArgThrSerIleGlyLysGln 375
779 AACCAATAGCGAATGTGTCGACCAAAATGTTCAAGACCAAGGATTC 828
|||
375 alPheLysValGlnIleLeu.....AsnCysThrGlnAsp..... 386
829 GGAGCTGTTAAACAGAGTGTCGCCGACCAAACTGCTGACAGTCGCGCA 878
|||
387IleLysAlaArgLeuIleAlaGlnLysProLeuSerLysTyrVa 401
879 ACTTGT..... 885
401 lleuArgLysLeuGlnAlaCysGlnHisProIleGlyGlnTyrCysMet 418
886TTACTCAAGAAAGATCTGCA.....GAACCG 912
|||
418 eTyrProLysLeuLeuIleLysLysAsnSerAlaThrGlnIleGlnGlu 434
913 GAGATATATGATGTACGAACGTATATCAACACCCCTTGAATTAAGCA 962
|||
435 GluAsnLeuCysAspSerLeuLeuLysAsnGlnGlnAlaIaLysGln 451
963 TGATATCAAGCTTTCAGCTGAT 987
451 yGlnAsnLysCysValLysValAsp 459
seq_name: /cgn2_6/ptodata/1/lae/5A_COMB.pep:US-08-325-071-61
seg_documentation_block:
; Sequence 61, Application US/08325071
; Patent No. 5587311

GENERAL INFORMATION:
APPLICANT: COBON, Stewart Gary
APPLICANT: MOORE, Joanna Terry
APPLICANT: JOHNSON, Law Anthony York
APPLICANT: WILADSEN, Peter
APPLICANT: KEMP, David Harold
APPLICANT: SRISKANTHA, Alagacoe
APPLICANT: RIDING, George Alfred
APPLICANT: RAND, Keith No. 5587311man
TITLE OF INVENTION: DNA Encoding A Cell Membrane
TITLE OF INVENTION: Glycoprotein of A Tick Gut
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/325,071
FILING DATE: 14-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,109
FILING DATE: 17-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/926,368
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/242,196
FILING DATE: 06-JUL-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/A087/00401
FILING DATE: 27-NOV-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P14912
FILING DATE: 16-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P12570
FILING DATE: 19-JUN-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PH9196
FILING DATE: 27-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 60042/111 BIAU
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 549 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-325-071-61
alignment_scores:
Quality: 89.00 Length: 359
Ratio: 0.614 Gaps: 22
Percent Similarity: 40.390 Percent Identity: 22.284
alignment_block:
US-09-323-427-3 x US-08-325-071-61 ..
Align seg 1/1 to: US-08-325-071-61 from: 1 to: 549

138 TGCATCGAAGGACATGT.....TTATG 160
|||||
66 CysValProThrThrcysLeuArgProAspLeuThcysLysAspLeu 82
161 TGAAGGCTTTATGATCAAGAAGGTGGCGTAATGATGAGGTGACG 210
|||||
82 sgluylsasleuleuglnargasp.....Sera 92
211 CAAGTTCGCGGAATTCACCTTCATTCATGCAATGTCGGGTAC 260
|||||
92 rgscyscysglnlytrpasnthr.....AlasncysSerAlaIa 105
261 ACATCTCTGATCCACGCGTATTTTGTACAC..... 296
|||||
106 ProProAlaAspser.....TyrCysSerProGlySerProLysgl 119
297ACTGTGTCAT.....TTCG 312
119 yProAspGlyGlnCysIleAsnAlaCysLysMetLysGluAlaGlyPheV 136
313 TTTCATCCATTTATTTGTTCACCAAGTTGATCGTATTCGATGACAAAG 362
|||
136 alCysgluHnIsglyCys.ArgSerThrAlaLysAlaLysGlnCysThrcy 152
363 C.....TTTACATGGAACCGTATTAACAGTTAGTGCACAGATG 403
152 sProArgGlyPheThrValAlaGluAspGlyLleThrCysLysSerIle 169
404 AGGTATCTGAATCACAACCTGCTTTCAACATCAATTC.....CCGANG 450
|||
169 erHisThrValserCysThrAlaGluGlnLysGlnThrCysArgProThr 185
451 CCAGTATGCCGTTATGAA.....ATTTGAGAGGTGGACCAACCG 491
186 GluAspCysArgValHisLysGlyThrValLeuCysGlnCysProTrpAs 202
492 TCACACAGTTCAATTTGCTATCATGTCGACGCCAGTTTATCAAAATGG 541
202 ngIn.....HisLeuValGlyasp.....T 209
542 CATGCGATTCGTAACCGTTGATACCTTTCGCGCGTTCGATTCCTGC 591
|||||
209 hrcysIleSerAspCysValAspLysLysCys.....HisGluGlu 222
592 TTGTGTCATGATGCT..... 606
223 PheMetAspCysGlyValTyrMetAsnArgGlnSerCysTyrCysProTr 239
607ACCGTGATACGTGGAAATTCAAATGCGATGAGAT 643
239 pLysSerArgLysProGlyProAsnVal.....AsnIleAsnGlyC 253
644 GTGCTCTGATAAATATTGCTAAAT.....AATTGGAA 678
|||
253 yslLeuLeuAsnGluTyrTyrThrValSerPheThrProAsnIleSer 269
679 TATCAACAGATTTAATGCTGCGCAAGAAGCTCAGTATACAAATATGC 728
|||
270 PheAspSerAsp.....HisCysLysTrpTyrGlu 279
729 GGATTCGATCACAGCTTTTCTATCAATGCCAGATCAGATACATTAAAG 778
|||||
279 uAspArg.....ValLeuGlnAlaIleArgTrpSerIleGlyLysGluV 294
779 AACCAATAGCAATGTGTTGACACCAATGTTACAGACCAAGATTC 828
|||
294 alPheLysValGluIleLeu.....AsnCysThrGlnAsp..... 305
829 GGAGCTGTTAAACAGGTGTCGCCGACGAAACCTGCTGCACCTCGCA 878
|||
306IleLysAlaArgLeuIleAlaGluLysProLeuSerAsnHisVa 320

```

879 ACTTCGT..... 885
      |||||
320 lLeuArgLysLeuGlnAlaCysGluHisProIleGlyGluTrpCysMetM 337
886 .....TTACAGAGAAAGATCTCA.....GAGCCG 912
337 eLtyrProLysLeuLeuIleLysLysAsnSerAlaTrpGluIleGlu 353
913 GAGAAATATCATGTATGATACGAACTGATATCAACACCTTGAAATACCGA 962
354 GluAsnLeuCysAspSerLeuLeuLysAsnGlnGlnAlaIleTyrLysG1 370
      |||||
963 TGATATCAAGCTTTGCCACTTGAT 987
      |||||
370 yGlnAsnLysCysValLysValasp 378

seq_name: /cgn2_6/ptodata/1/laa/5A_COMB.pep:us-08-325-071-63

seq_documentation_block:
; Sequence 63, Application US/08325071
; Patent No. 5587311
; GENERAL INFORMATION:
; APPLICANT: COBON, Stewart Gary
; APPLICANT: MOORE, Joanna Terry
; APPLICANT: JOHNSON, Law Anthony York
; APPLICANT: WILLADSEN, Peter
; APPLICANT: KEMP, David Harold
; APPLICANT: SRISKANTHA, Alagacoe
; APPLICANT: RIDING, George Alfred
; APPLICANT: RAND, Keith No. 5587311man
; TITLE OF INVENTION: DNA Encoding A Cell Membrane
; TITLE OF INVENTION: Glycoprotein Of A Tick Gut
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325, 071
; FILING DATE: 14-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/062,109
; FILING DATE: 17-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,368
; FILING DATE: 07-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/242,196
; FILING DATE: 06-JUL-1988
; PRIOR APPLICATION DATA:
; APPLICATION NMBER: PCT/AU87/00401
; FILING DATE: 27-NOV-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU P14912
; FILING DATE: 16-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU P12570
; FILING DATE: 19-JUN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PH9196
; FILING DATE: 27-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 60042/111 B1AU
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 650 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-325-071-63

alignment_scores:
      Quality: 89.00      Length: 359
      Ratio: 0.614      Gaps: 22
Percent Similarity: 40.390      Percent Identity: 22.284

alignment_block:
US-09-323-427-3 x US-08-325-071-63 ..
Align seq 1/1 to: US-08-325-071-63 from: 1 to: 650

138 TGCATTCGAAAGACATGT.....TTATG 160
      |||:|:| |||||
147 CysValProThrThrcysLeuArgProAspLeuThrcysLysAspLeu 163
161 TGAAGGCTTTTATGATCAAGAGAGTTGCCGTATGATGAGTGAGCGT 210
      |||:|:| |||:|:|
163 sGluLysAsnLeuLeuGlnArgasp.....Sera 173
211 CAAGTTCGCGGAATTCACATTCATTTGATTCGAAATGCGGCGTAC 260
      |||:|:|:|:|:|:|:|
173 rGcysCysGlnGlyTrpAsnThr.....AlaAsnCysSeraAlaIa 186
261 ACATCTCTGATCAATCAACAGTGATTTTGTAAACAC..... 296
187 ProProAlaAspSer.....TyrCysSerProGlySerProLysG1 200
297 .....AACTGTGTGAT.....TTGC 312
200 yProAspGlyGlnCysIleAsnAlaCysLysMetLysGlnAlaGlyPheV 217
313 TTTCATTCATTTATTTGTACCAAGTTGATCGTGATATGAGTACAAATG 362
      ||:|:| |||:|:|:|:|:|:|
217 aLcysGlnHisGlyCys.ArgSerThrAlaLysAlaTyrGluCysThrCy 233
363 C.....TTTACATGAGACGTGATTAACAGCATGAGCAGATG 403
233 sProArgGlyPheThrValAlaGlnAspGlyIleThrCysLysSerIleS 250
404 AGGTATCTGAATACACACTGCTTTTCAACATCAATTGTC...CCGATG 450
      ||:|:| ||||| ||| |||
250 eHisThrValSerCysThrAlaGlnGlnLysGlnThrCysArgProThr 266
451 CCAGTATGCCGTTATGAA.....ATTTGATGATGGGACCAACCGG 491
267 GluAspCysArgValHisLysGlyThrValLeuCysGluCysProTrpAs 283
492 TCACACAGTTCAATTTGCTATCATTTGTCACGACGTTATCATTAATGA 541
283 nGln.....HisLeuValGlyasp.....T 290
542 CATGCGATTCGAAACCGTTGATACTTTGCGCGGTTGTCATTCCTGCG 591
      ||| |||:|:| ||||| |||
290 hCysIleSerAspCysValAspLysLysCys.....HisGluGlu 303
592 TTTGTCGATGATGT..... 606
304 PheMetAspCysGlyValTyrMetAsnArgLysCysTyrCysProTr 320
607 .....AACGGTATACGTGGAATTCGAATTCGAATGCGTATGATGAT 643
320 pLysSerArgLysProGlyProAsnVal.....AsnIleAsnGlyC 334

```

```

644 GTGCTTGATAAATATTTGCTAAT.....AATTGAA 678
   || |||:||||| :|||:|||||
334 yLeuLeuAsnGluTyrTyrThrValSerPheThrProAsnIleSer 350
679 TATCCAAACAGATTTAATGGCTGCCAAGAAGCTACAGTATACAAATATGC 728
   :|||:||||| :|||:|||||
351 PheAspSerAsp.....HisCysLysTrpTyrG1 360
729 GGATGATGACACGCTTTTCTATCAATGCACAGATCATATTACATTAAAG 778
   |||||:||||| :|||:|||||
360 uAspArg.....ValLeuGluAlaIleArgThrSerIleGlyLysGluV 375
779 AACCAAAATGCAATGCTTTCGACCAACATGTCACAGACACACAGATTC 828
   :|||:||||| :|||:|||||
375 aLpHeLysValGluIleLeu.....AsnCysThrGlnAsp..... 386
829 GGAGCTGTTAAACAGGTGTGTCGCCGACGACAAAACCTGTGACGTCGCCGA 878
   :|||:||||| :|||:|||||
387 .....IleLysAlaArgLeuIleAlaGluLysProLeuSerAsnHisVa 401
879 ACTTGT..... 885
401 lleuArgLysLeuGlnAlaCysGluHisProIleGlyGluTrpCysMetM 418
886 .....TTACTCAAGAAAAGATCTGCA.....GAACCG 912
   |||||:||||| :|||:|||||
418 eTyrProLysLeuLeuIleLysLysAsnSerAlaThrGluIleGluG1 434
913 GAGCATATCATTTGATGTACGAACGATATCAACACCTTGAATATGACGA 962
   |||||:||||| :|||:|||||
435 GluAsnLeuCysAspSerLeuLeuLysAsnGlnGluAlaAlaTyrLysG1 451
963 TGATATATCAAGCTTTCGCACTGAT 987
451 yGlnAsnLysCysValLysValAsp 459

seq_name: /cgn2_6/ptodata/1/1aa/5a_COMB.pep:US-08-450-332-2
seq_documentation_block:
: Sequence 2, Application US/08450332
: Patent No. 5744350
: GENERAL INFORMATION:
: APPLICANT: VINCI, VICTOR A.
: APPLICANT: CONDER, MICHAEL J.
: APPLICANT: MCADA, PHYLLIS C.
: APPLICANT: REEVES, CHRISTOPHER D.
: APPLICANT: DAVIS, CHARLES R.
: APPLICANT: HENDRICKSON, LEE E.
: APPLICANT: RAMBOSEK, JOHN
: TITLE OF INVENTION: DNA ENCODING TRIOL POLYKETIDE SYNTHASE
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CHRISTINE E. CARTY
: STREET: 126 E. LINCOLN AVENUE, P. O. Box 2000
: CITY: RAHWAY
: STATE: NJ
: COUNTRY: USA
: ZIP: 07065-0907
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/450,332
: FILING DATE: 25-MAY-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: CARTY, CHRISTINE E.
: REGISTRATION NUMBER: 36,099
: REFERENCE/DOCKET NUMBER: 19076CA
: TELECOMMUNICATION INFORMATION:

```

```

: TELEPHONE: 908-594-6734
: TELEFAX: 908-594-4720
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3038 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: YES
: ORIGINAL SOURCE:
: ORGANISM: TPMS Protein
: US-08-450-332-2

alignment_scores:
  Quality: 87.00      Length: 472
  Ratio: 0.481      Gaps: 24
  Percent Similarity: 38.347      Percent Identity: 19.703

alignment_block:
US-09-323-427-3 x US-08-450-332-2 ..

Align seg 1/1 to: US-08-450-332-2 from: 1 to: 3038

16 ATTGCTTTCTGACTACACTTAATTCATGCTTTATTCGATTCGGTTGA 65
   |||||:||||| :|||:|||||
1175 ValAlaPheGlnThrValIleGlyAla.....TyrSerSerPro..... 1187
66 CAATGCTGTGGAAGTGACCGACGACAAATTTGAATGT.....GGACCAACTT 109
   |||||:||||| :|||:|||||
1188 .....GlyAspArgArgLeuArgCysLeuTyrValProThrH 1200
110 CAATA..... 114
1200 isValAspArgIleThrLeuValProSerLeuCysLeuAlaThrAlaGlu 1216
115 .....ACATCAATTTTAAATACAGTAATGCAATTCGAAGACAGA 152
1217 SerGlyCysGluLysValAlaPheAsnThrIleAsnThrTyrAspLysG1 1233
   |||||:||||| :|||:|||||
1233 yAspTyrLeuSerGlyAspIleValAlaPheAspAlaGlu..... 1246
194 ATGATCAAGCTGACGCTCAAGTTGCCGGAATTCACCTT...CCATTGAT 240
   :|||:||||| :|||:|||||
1247 ..GlnThrThrLeuPheGlnValGluAsnIleThrPheLysProPheSer 1262
241 TCATGCAATGTTGGCGGTACAGCATCTGCAATCCAGTGAATTTT... 288
1263 ProProAspAlaSerThrAspHisAlaMetPheAlaArgTrpSerTrpG1 1279
289 ...GTACCAACAAGCTGTTCATTTGCTTTCATCCATATTATTTGTTACA 334
   :|||:||||| :|||:|||||
1279 yProLeuThrProAspSerLeuLeuAspAsnProGluTyrTrpAlaThrA 1296
335 AAGTGTATGCT.....GCATATCGAGTACAAATGCTTTTTC 369
   |||||:||||| :|||:|||||
1296 laGlnAspLysGluAlaIleProIleIleGluArgIleValTyrPheTyr 1312
370 ATGGAAGCTGATAAACAAGCTTAGTGCACAGATTGAGGTTCGGAATTCAC 419
   :|||:||||| :|||:|||||
1313 Ile.....ArgSerPheLeuSerGlnLeuThrPheLysGluGluArgG1 1326
420 AACTGCT.....TTTCAACTCAAAATGTTCGCGATGCCAGTATGCC 460
   |||||:||||| :|||:|||||
1326 ngIlnAlaIlePheHisLysGlnLysGlnIleGluTrpLeuGlu..... 1340
461 GTTATGAATTTTGGATGGTGGACCAACCGGTCAACCAAGTCAATTTGCT 510
   :|||:||||| :|||:|||||
1341 .....GlnValLeuAlaSerAlaLysGluGluArg..... 1350

```

```

511 ATCATGTGTAGCAGCACTTTATCATTAATTAAGTGCAGCATGGCAATGGCAACCGT 560
1351 .....HisLeuTrrTyrAspProGlyTrpGluAs 1360
561 TGAATACTTTCGCGCGGTGTCCATTCCTCCCTGTTCGATGAT..... 603
1360 nAspTrhGluIaGlnIleGluHisLeuCysTrhAlaAsnSerTyrHisP 1377
603 ..... 603
1377 roHISValArgLeuValGlnArgValGlnHisLeuLeuProThVal 1393
604 ...GGTAACGGTGTACTGTGAAATTTCAATGCTGATGATGGTGCCT 650
1394 ArgSerAsnGlyAsnProPheAspLeuAspHisAspGly...LeuLe 1409
651 TGAATAATATTTCGTAATAT.....TTGGAATATC 682
1409 uTrhGluPheTyrThrAsnThrLeuSerPheGlyProAlaLeuHisTyrA 1426
683 CAACAGATTAAATGGCGGCGCAAGCAACCTCAGATTACAATATAGCGAT 732
1426 IaArgLeuLeuValAla...GlnIleIaHisArgTyrGlnSerMetAsp 1441
733 CGATTCACAGCTTTTCTATCAATGCGACGATAGTATTACATTAAAGAAC 782
1442 IleLeuGlnIle..... 1445
783 AATATAGCAATGTGTTCGACCACATGTTCAGAACCAAGATTCGGAG 832
1446 .....GlyAlaGly. 1448
833 CTGTAAACAGGTGTGTGCGGACAGCAACCTGCGAGCTGCGCAACT 882
1449 .....ThrGlyGlyAlaThrTyrGlyValLeuAlaThrProGlnLeu 1462
882 ..... 882
1463 GlyPheAsnSerTyrThrTyrThrAspIleSerThrGlyPheGluG 1479
882 ..... 882
1479 nAlaArgGluGlnPheAlaArgPheGluAspArgMetValPheGluPro 1496
883 .....CGTTTACTCAGAAAAAGATCTGCAGACCGGAGATATATC 921
1496 euAspIleArgArgSerProAlaGlnIleGlyPheGluPro...HisAla 1511
922 ATTAGAGTACAGATCGATATACACACCCCTGAAATTAGCATGATATCA 971
1512 TyrAspLeuIleIleAlaSerAsnValLeuHisAlaThrProAsp... 1526
972 AGCTTTCGCAAGTTGATTTACGTAC...CGTGACATTCCTGCACAAATATG 1018
1527 ...LeuGlnTyrThrMetAlaHisAlaArgSerLeuLeuLysProGly 1542
1019 GACACACTGTAAATTT..... 1035
1542 IylnMetValIleLeuGluIleThrHisLysGlnHisTrhArgLeuGly 1558
1036 .....GTGCACTACAAATATGCAAT 1055
1559 PheIlePheGlyLeuPheAlaAspTrpTrpAlaGlyValAspArgGlyAr 1575
1056 CTGCATGTCCACCATTT 1071
1575 gCysTrhGluProPhe 1580
seq_name: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:us-08-637-640-2
seq_documentation_block:
; Sequence 2, Application US/08637640
; Patent No. 5849541

```

```

GENERAL INFORMATION:
APPLICANT: VINCI, VICTOR A.
APPLICANT: CONDER, MICHAEL J.
APPLICANT: MCADA, PHYLLIS C.
APPLICANT: REEVES, CHRISTOPHER D.
APPLICANT: DAVIS, CHARLES R.
APPLICANT: HENDRICKSON, LEE E.
APPLICANT: RAMBOSEK, JOHN
TITLE OF INVENTION: DNA ENCODING TRIOL POLYKETIDE SYNTHASE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESSES:
ADDRESSER: CHRISTINE E. CARTY
STREET: P. O. BOX 2000; 126 E. LINCOLN AVENUE
CITY: RAHWAY
STATE: NJ
COUNTRY: USA
ZIP: 07065

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/637,640
FILING DATE: 23-AUG-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/148,132
FILING DATE: 01-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: CARTY, CHRISTINE E.
REGISTRATION NUMBER: 36,099
REFERENCE/DOCKET NUMBER: 19076
TELECOMMUNICATION INFORMATION:
TELEPHONE: 908-594-6734
TELEFAX: 908-594-6720
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3038 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEetical: YES
ORIGINAL SOURCE:
ORGANISM: TPKS Protein
US-08-637-640-2

alignment_scores:
Quality: 87.00 Length: 472
Ratio: 0.481 Gaps: 24
Percent Similarity: 38.347 Percent Identity: 19.703

Alignment_block:
US-09-323-427-3 x US-08-637-640-2 ..

Align seg 1/1 to: US-08-637-640-2 from: 1 to: 3038

16 ATTGCTTTCTGACTACACTTATTCATTGCTTATTCGATTCGGGTGA 65
:::||||| |||::: ||| ||||| |||
1175 ValAlapheGlnThrValIleGlyAla.....TyrSerSerPro..... 1187
66 CAATGCTGTGAGAGCGACCCAGAAATGATGT.....GGACCACACTT 109
||||: :::::||||| |||||
1188 .....GlyAspArgArgLeuArgCysLeuTyrValProThrH 1200
110 GATTA..... 114
:::
1200 IsValAspArgIleThrLeuValProSerLeuCysLeuAlaThrIaGlu 1216
115 .....ACAATCAATTTTATACACGTATGCAATTCGAGAGACA 152
:: ||||| |||:::

```



```

1217 SerGlyCysGlnLysValAlaPheAsnThrIleAsnThrTyrAspLysG1 1233
1218 TGTGTATGTCGAAGGT.....CTTATGATCAAGAAGTTGCCGTA 193
1219 yAspTyrLeuSerGlyAspIleValAlaPheAspAlaGlu..... 1246
1220 ATGATGAAGGTGAGCTCAAGTTGCCGAATTTACCT...CCATTGAT 240
1221 ..GlnThrThrIleuPheGlnValIleGlnAsnIleThrPheLysProPheSer 1262
1222 TCAATGCAAGTTGCGGTACACAGATCTGTGATCCACAGTGTATTTTT.. 288
1223 ProProAspLaserThrAspHisAlaMetPheAlaArgTrpSerTrpG1 1279
1224 ..GTACACAACTGTTGTCATTGCTTTCATCCATTATTGTTACCA 334
1225 yProLeuThrProAspSerLeuLysAspAsnProGluTyrTrpAlaThrA 1296
1226 AAGTTGATCGT.....GCATATGCAAGTACAAATGCTTTTAC 369
1227 lacIlnAspLysGluAlaIleProIleIleGluArgIleValTyrPheTyr 1312
1228 ATGGAAGCTGATAAACAGTTAGTCACAGATGAGGATGATCTGAATCAC 419
1229 Ile.....ArgSerPheLeuSerGlnLeuThrLeuGluIleArgG1 1326
1230 AACTGCT.....TTTCAACTCAAAATTGTCGCCAGTATGCC 460
1231 ngInAlaIleAlaPheHisLeuGlnLysGlnIleGluTyrLeuGlu..... 1340
1232 GTTATGAAATTTTGATGGTGGACCAACCGTCAACCGTTCAATTTGCT 510
1233 ..GlnValLeuAlaSerAlaLysGluLysArg..... 1350
1234 ATCATTTGGTCAGCCAGTTTATCATTAATGACATGGCATTTCTAAACCGT 560
1235 ..HisLeuTrpTyrAspProGluTyrPheGluAs 1360
1236 TGTATCTTTTGGCGGTTGTCATTCCTGCTTTCGATGATGAT..... 603
1237 nAspThrGlnAlaGlnIleGluHisLeuCysThrAlaAsnSerTyrHisP 1377
1238 .. 603
1239 roHisValArgLeuValGlnArgValGlyGlnHisLeuLeuProThrVal 1393
1240 ..GCTAACGGTACTACTGGAATTTCTAAATCGATGATGCTGCTCT 650
1241 ArgSerAsnGlyAsnProPheAspLeuAspHisAspGly...LeuLe 1409
1242 TGTAAATATTTTCTAAATAT.....TTGGAATATTC 682
1243 uThrGluPheTyrThrAsnThrLeuSerPheGlyProAlaLeuHisTyrA 1426
1244 CAACAGATTAAATGCTGGCCAAAGACTCAGTATACAAATATGCGGAT 732
1245 laArgGluLeuValAla...GlnIleAlaHisArgTyrGlnInsertMetAsp 1441
1246 CGATTCACAGCTTTTCTATTCATGTCAGATCAGTATTAACATTAAGAAC 782
1247 IleLeuGlnIle..... 1445
1248 AATATGCAATGTGTGCGACACATGTTCAAGAACCAAGATTTGGAG 832
1249 ..GlyAlaIleLys 1448
1250 CTGTAAACAGAGGTGGTGGCCGACAAACCTGCTGAGCTGGCACTT 882
1251 ThrGlyGlyAlaThrLysTyrValLeuAlaThrProGlnLeu 1462
1252 .. 882
1253 GlyPheAsnSerTyrThrTyrThrAspIleSerThrGlyPhePheGluG1 1479

```

```

882 ..... 882
1479 nAlaArgGlnGlnPheAlaProPheGlnAspArgMetValPheGluProL 1496
883 .....CGTTATCTCAAGAAAAGATCTGCAGAACCGAGAAATATTC 921
1496 euAspIleArgSerProAlaGlnGlnIlePheGluPro...HisAla 1511
922 ATGATGTLGACAGCTGATATCAACACCTTGAATTAAGCATGATATATCA 971
1512 TyrAspLeuIleIleAlaSerAsnValLeuHisAlaThrProAsp.... 1526
972 AGCTTGCSCAGTATTTACGTCAC...CGTGCATCTTGCACAAATATG 1018
1527 ..LeuGlnLysThrMetAlaHisAlaArgSerLeuLeuLysProGlyG 1542
1019 GACAACTGTAATATCTT..... 1035
1542 LysIleMetValIleLeuGluIleThrHisLysGlnHisThrArgLeuGly 1558
1036 .....GCTGCAGTACAAATGGAAT 1055
1559 PheIlePheGlyLeuPheAlaAspTrpTrpAlaGlyValAspAspGlyAr 1575
1056 CTCATGTCACCAATTT 1071
1575 gCysThrGluProPhe 1580

seq_name: /cgn2.6/plodata/1/1aa/5B.COMB.pep:US-08-540-804-12
seq_documentation_block:
: Sequence 12, Application US/08540804
: Patent No. 5919666
: GENERAL INFORMATION:
: APPLICANT: Young, Richard A.
: APPLICANT: Koleske, Anthony J.
: APPLICANT: Thompson, Craig M.
: APPLICANT: Chao, David M.
: TITLE OF INVENTION: No. 5919666el Factors Which Modify Gene
: TITLE OF INVENTION: Transcription and Methods of Use Therefor
: NUMBER OF SEQUENCES: 39
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
: STREET: Two Militia Drive
: CITY: Lexington
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02173
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/540,804
: FILING DATE: 11-OCT-1995
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/521,872
: FILING DATE: 21-AUG-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/218,265
: FILING DATE: 25-MAR-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Granahan, Patricia
: REGISTRATION NUMBER: 32,227
: REFERENCE/DOCKET NUMBER: WH194-03A2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-861-6240
: TELEFAX: 617-861-9540
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 1226 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-540-804-12

```

```

alignment_scores:
  Quality: 85.50      Length: 390
  Ratio: 0.457       Gaps: 13
Percent Similarity: 47.949   Percent Identity: 19.744

```

```

alignment_block:
US-09-323-427-3 x US-08-540-804-12

```

```

Align seg 1/1 to: US-08-540-804-12 from: 1 to: 1226

```

```

14  TTAATGCTTTCTGACTACACTTATTCGATTGCTTATTCGATTCCGGTT 63
    ::::: |||::: ::::: |||::: |||::: |||:::
358 MetMetProSerLeuTyrArgLeuLeuAsnIleLeuIleThrTyrGlyI 374
    : GACAAATGCTGCGAAGGTGAGCCGAAATGGAATGTGACCAACTTCAAT 113
    : |||::: |||::: |||::: |||:::
374 eIleIysValProThrTyrIleArgIysLeuIle..... 385
114 AACAAATCAATTTTAATACAGTAATGCAATTCGAAGACATGTTATGTA 163
    ::::: |||::: |||::: |||::: |||::: |||:::
386 ..SerSerGlyLeuLeuTyrLeuGlnAspSerAsnAspIysPheValHis 401
164 AAGCTCTTTATGATCAG...AAGGTGCGCGTATGATGAAAGTGAGAGCT 210
    ::::: |||::: |||::: |||::: |||::: |||:::
402 ValGlnLeuLeuIleAsnLeuLysIleSerProLeuMetLys..... 415
211 CAAGTTCGCGAATTCCTTCATTCATTCATTCATTCATTCATTCATTCAT 260
    ::::: |||::: |||::: |||::: |||::: |||:::
416 .....SerGlnTyrAsnMetValIleArgAsnValMetLut 428
261 ACGATCTCTGAATCCACGCTGATTTTGTACACACACTG..... 301
    ::::: |||::: |||::: |||::: |||::: |||:::
428 YrsPValIysPheTyrGluIlePheAsnAspGlnLeuValGluIle 444
302 .....TGTCTCATTTGCTTTCATTCATTCATTCATTCATTCATTCATTC 327
445 ThrGlnIleIleLysMetArgIleLeuSerAsnAspIleThrAsnLeuG 461
328 GTTACCAAGCTGATCGCATATCGATACGAATGCTTTTACATGGAAGC 377
    |||::: |||::: |||::: |||::: |||::: |||:::
461 nLeuSerLysThrProLeuSerIleLysIleMetValAlaGluTyrIle 478
378 TG.....ATTAACAGTTAGTGCACAGATTGAGGTAT 409
    |||::: |||::: |||::: |||::: |||::: |||:::
478 euserHisLeuCysserGlyIleLeuSerSerValAsnArgThrValLeu 494
410 CTGAATACCAACTGCTTTC.....AACTCAATATGTCGCGATCCCA 453
    |||::: |||::: |||::: |||::: |||::: |||:::
495 LeuLysIlePheLysIlePheCysIleAspLeuGlnValAlaPheHisSph 511
511 ePheLysTrpIleGlnPheIleValIleValIleGlnLeuLeuSerAspIleG 528
504 ATTTGCTA.....TCATTGCTC 520
    ::::: |||::: |||::: |||::: |||::: |||:::
528 IserLeuGlnAlaIleuMetAspIleLeuLeuLysTyrGlnLysLeuPhe 544
521 AGCCAGTTTATCATTAATGACATGCAATTCG...AAACCGTTGATA.. 565
    |||::: |||::: |||::: |||::: |||::: |||:::
545 SerGlnPhe...IleAsnAspHisIleLeuPheThrLysThrPheIlePhe 560
566 .....CTTTCGCGCGGTGTCATTCCTCTGTTGCGATGA 602
    ::::: |||::: |||::: |||::: |||::: |||:::
560 eIleTyrLysLysValLeuLysGlnLysAspValProAlaTyrAsnValT 577

```

```

603 TGGTAACGCTGATACGTGTGAAATTTCTAAATGCTGATGATGTCCTTGG 652
    ::::: |||::: |||::: |||::: |||:::
577 hrserPheMetProPheTrpLysPhePheMetLysAsnPheProPheVal 593
653 ATAAATTTTGTGTAATTAATTTGGAATATCCACACAGTTTAATGGCGGC 702
    ::::: |||::: |||::: |||::: |||:::
594 LeuLysValAspAsnAspLeuArgIleGlnLeuGlnSerValIleTyrAsn 610
703 CAAGAACCTCAGCTATACAAATATGCGGATGCAATCAGCTTTTCTATCA 752
    ::::: |||::: |||::: |||::: |||::: |||:::
610 pGluLysLeuLysThrGlnLysLeuLysAsnAspLysSerGlnValLeuL 627
753 ATGCCAGATCAGTATTACCATTAAGAACCA..... 784
    ::::: |||::: |||::: |||::: |||::: |||:::
627 yValIleTyrSerMetIleAsnAsnSerAsnGlnAlaValGlyIleThrTyr 643
785 ..ATAGCGAATGTGTGCGACCAACATGT..... 811
    ::::: |||::: |||::: |||::: |||::: |||:::
644 AsnPheProGlnValPheGlnValAsnIleArgPheLeuLeuHisAsnSe 660
812 .....CAGAACCAACAGGATT 827
    ::::: |||::: |||::: |||::: |||::: |||:::
660 rGluIleIleAspThrAsnThrSerLysGlnPheGlnLysAlaIleArgAsn 677
828 CGGAGCTGTTAAACAGAGTGTCGCCGCAAAACCTGTCAGCTGCGC 877
    ::::: |||::: |||::: |||::: |||::: |||:::
677 snValMetLeuLeuIleAlaThrAsnLeuLysLysLysIleLysPheMet 693
878 AACTGCTTACCTCAGAAAGATCTGCAGACCGCGGAGATA..... 919
    ::::: |||::: |||::: |||::: |||::: |||:::
694 SerIlePheLeuLysArgLysAspPheThrAsnLysAsnLeuIleGln 710
920 .....TCATGTATGATGACGATGATGATGATGATGATGATGATGAT 956
    ::::: |||::: |||::: |||::: |||::: |||:::
710 uIleSerLeuLysLeuLeuThrPheGlnValAlaThrGlnAsnValLeuGly 727
957 TAGCGATGATATCAGCTT 976
    |||::: |||::: |||::: |||::: |||::: |||:::
727 eugLutTyrIleIleArgLeu 733

```

```

seq_name: /cgn2_6/ptodata/1/laa/5B_COMB.pep:US-08-218-265-12

```

```

seq_documentation_block:

```

```

; Sequence 12, Application US/08218265
; Patent No. 5922585
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A.
; APPLICANT: Young, Richard A.
; APPLICANT: Thompson, Craig M.
; TITLE OF INVENTION: No. 5922585el Factors Which Modify Gene
; TITLE OF INVENTION: Transcription and Methods of Use Thereof
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millita Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: US
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/218,265
; FILING DATE: 25-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI94-03
; TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: 617-861-6240
 ; TELEFAX: 617-861-9540
 ; INFORMATION FOR SEQ ID NO: 12:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1226 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-218-265-12

alignment_scores:
 Quality: 85.50 Length: 390
 Ratio: 0.457 Gaps: 13
 Percent Similarity: 47.949 Percent Identity: 19.744

alignment_block:
US-09-323-427-3 x US-08-218-265-12 ..

Align seg 1/1 to: US-08-218-265-12 from: 1 to: 1226

```

14  TTTATGCTTCTGTACTACACTATTCATGCTCTATTCGATCCGGTT 63
    ::::: |||::: ::|||::: |||::: :::
358 MetMetProSerLeuYrArgLeuAsnIleLeuIleThrYrGluI 374
    : |||
64  GACAAATGCTCGAAGTGCAGCAAAATTGAATGTGACCAACTTCAAT 113
    : ||| ::|||:::
374 eileYsValProThrYrIleArgYsLeuIle..... 385
    : |||
114 AACATCAATTTTAAATACAGTAATGCATTCGAGACATGTTATGTGA 163
    : ||| ::|||::: |||::: |||:::
386 ..SerSerGlyLeuLeuYrLeuGlnAspSerAsnSplysPheValHis 401
164 AAGCTTTATATGATCAAG..AAGTGGCCGTAATGATGAGGTGACGT 210
    : |||::: |||::: |||::: |||:::
402 ValGlnLeuLeuIleAsnLeuYsIleSerProLeuMetLys..... 415
211 CAAGTTGCGGAATTCACCTTCATTTGATTCATGCAATGTTCGCGTAC 260
    : |||::: |||::: |||::: |||:::
416 .....SerGlnYrAsnMetValIleuArgAsnValMetGluTr 428
261 AGCATCTCGAATCCAGTGTGATTTTGTACACAACTG..... 301
    : |||::: |||::: |||::: |||:::
428 YrAspValIlyspPheYrGluIlePheAsnPhaspGlnLeuValGluIle 444
302 .....TTGTCAATTTCGTTTCATTCATCATATT 327
445 ThrGluGlnIleYsMetArgIleLeuSerAsnAspIleThrAsnLeuG 461
328 GTTACCAAGTTGATCGATCGATTCGATCAATGCTTTTACATGGAAGC 377
    : ||| ||| ::|||::: |||::: |||:::
461 nleuSerIlyThrProLeuSerIleYsIleMetValAlaGluTrpYrL 478
378 YG.....ATTAACAGTTAGTGCACAGATTCAGAGTAT 409
    : ||| ||| ::|||::: |||::: |||:::
478 euserHisLeuYsSerGlyIleLeuSerSerValAsnArgThrValLeu 494
410 CTTGAATCAACAACCTGCTTTC.....AAACTCAATTTGCCGAGGCCA 453
    : ||| ||| ::|||::: |||::: |||:::
453 LeuYsIlePheYsIlePheYsIleAspLeuGlnValPheHisHisPhe 511
495 GATAGCGGTTATGAAATTTGGATGGTGACCAACCGGTCAACAGTTCA 503
    : ||| ||| ::|||::: |||::: |||:::
511 ePheYsTrpIleGluPheIleValYrHisGlnLeuLeuSerAspIleG 528
504 ATTTCGTA.....TCATTGGTC 520
    : ||| ||| ::|||::: |||::: |||:::
528 lueSerLeuGlnAlaLeuMetAspIleLeuLeuYsTrpGlnIlySleuPhe 544
521 AGCAGGTTTATCATTAATGAGCATGGATTCTG..AAACCGTTGATA.. 565
    : ||| ||| ||| ||| ::|||::: |||::: |||:::
545 SerGlnPhe...IleAsnAspHisIleLeuPheThrIlyThrIlePhe 560
  
```

```

566 .....CTTTCGCGCGGTTCATTCCTGCTTGTTCGATGA 602
    : ||| ::|||::: |||::: |||:::
560 eileYrIlySlyValLeuYsGluYsAspValProAlaYrAsnValTr 577
    : ||| ||| ::|||::: |||::: |||:::
603 TGCTAACGCTGATCTGTGAAATTTCTAAATGCTGATGATGCTGCTTG 652
    : ||| ||| ::|||::: |||::: |||:::
577 hrSerPheMetProPheTrpYrPhePheMetLysAsnPhaPheVal 593
    : ||| ||| ::|||::: |||::: |||:::
653 ATTAATATTTCCTAATAATTGGAATATCCACACAGATTAAATGCGTGGC 702
    : ||| ||| ::|||::: |||::: |||:::
594 LeuYsValAspAsnAspLeuArgIleGluLeuGlnSerValYrAsnAs 610
703 CAAAGACTCACCCTAATACAAATATGCGGATCGATCAGCTTTTCTATCA 752
    : ||| ||| ::|||::: |||::: |||:::
610 pGluYsLeuYsThrGluYsLeuYsAsnAspYsSerGluValLeuL 627
753 ATCCAGATCAGTATTACATTAAGAACCA..... 784
    : ||| ||| ::|||::: |||::: |||:::
627 YsValYrSerMetIleAsnAsnSerAsnGlnAlaValGlyGlnThrTrp 643
785 ..ATAGCGAATGTTTCGACCAACATGTT..... 811
    : ||| ||| ::|||::: |||::: |||:::
644 AsnPheProGlnValPheGlnValAsnIleArgPheLeuLeuHisAsnSe 660
812 .....CAGAACACAGAGATT 827
    : ||| ||| ::|||::: |||::: |||:::
660 rGluIleIleAspThrAsnThrSerIlySglnPheGlnIlyAlaArgAsnA 677
828 CGGAGCTGTTAAACAGGTGTGCGCCGCAAAACCTGCTGACGTGCGC 877
    : ||| ||| ::|||::: |||::: |||:::
677 snValMetLeuLeuIleAlaThrAsnLeuYsGluYrIleYsPheMet 693
878 AACTGCTTACTCAGAAAGATCTGCAGAACCGAGATA..... 919
    : ||| ||| ::|||::: |||::: |||:::
694 SerIlePheLeuYsArgYsAspPheThrAsnIlyAsnLeuIleGlnLe 710
920 .....TCATTGATGACGACGATATACACACCTTGAAT 956
    : ||| ||| ::|||::: |||::: |||:::
710 uIleSerLeuYsLeuLeuThrPheGlnValThrGlnAsnValLeuGlyL 727
957 TAGCATGATAATCAAGCTT 976
    : ||| ||| ::|||::: |||::: |||:::
727 euGluYrIleIleArgLeu 733
seq_name: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:US-08-480-604A-6
seq_documentation_block:
; Sequence 6, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESS: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
  
```

```

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPND-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-604A-6

alignment_scores:
Quality: 85.00 Length: 196
Ratio: 0.825 Gaps: 9
Percent Similarity: 52.551 Percent Identity: 22.449

alignment_block:
US-09-323-427-3 x US-08-480-604A-6 ..

Align seg 1/1 to: US-08-480-604A-6 from: 1 to: 2710

529 TATCATTAATGACATGCGATTCGAAACGGTATCTTCTGCGGGGT 578
||||:|||||
188 TTTTTLTYSSERGILNLSNLYSPROTHVALPROTHRLLEASPAPII 204
||||:|||||
579 TGTSCATTCCTGCTTGTGCGATGATGGTAAACGGTATCTGTGAAATTC 628
||||:|||||
204 EILELVSSEHISLEUVALSERGLU.....T 213
629 TAAATGCTGATGATGCTCTTGATTAATATTGCTAAATATTGGAA 678
||||:|||||
213 YTASnTgrAspRgIuThValLeuGlUsErTyTgrThraSnSerLeuArg 229
679 TATSCAACAATTTAATGGCTGGCAAGAAGCTCACATTAACAATATGC 728
||||:|||||
230 ..Lys..IleA 232
729 GGATGCGATCACAGCTTTTCTATCATGCCAGATCAGTATTACCATTAAG 778
||||:|||||
232 snSerLsnHISGLYIleAspRIleArgIalnsnSerLeuPhe...ThGlu 247
779 AACCAATATGACGAATGTTTGCACACATGTTTCAGAACCAAGCATTC 828
||||:|||||
248 GInGlUleuLeuSnIleTySerGInGlUleuLeuSnArgGlysnIle 264
829 GGAGCGTGTAAACAGGTGCTGCCGCAAAACCTGCTGCAGCTGGCA 878
||||:|||||
264 uAlaIaIaIaSerAspIleVal.....ArgLeuLeuAlaIeLeuYsa 278

```

```

879 ACTTC.....GTTTACTCAAGAAAGATCTGCAGAACCGGAGAAATATCA 922
|||||
278 snPhleGlyGlyValTYrLeuAspValaspMetLeuProGlyIleHisSer 294
|||||
923 TTGATGTACGAAGTATATCAACACCC..... 949
|||||
295 AspLeuPhleYThrIleSerArgProSerSerIleGlyLeuAspArgTr 311
950 .....TGAATTTAGCGATGATTAATCAAGCTTTGCCAGTTGATT... 988
311 pGluMetIleLeuSerGluAlaIleMetIleTYrLYsTYrIleAsnA 328
989 ..TACGTACACCGTGCACTTC.....TGCAACATTAATGACACAC 1024
|||||
328 snTYrThrSerGluAsnPhleAspLYsLeuAspGInGInLeuLYsAspAsn 344
1025 CTGTAATACTTGCTGCAGTACAAATGGAATCTGCA 1060
|||
345 PhelysLeuIleIleGlySerLYsSerGluLYsSer 356

seq_name: /cgn2_6/plodata/1/1aa/5B_COMB.pep:US-08-405-496A-6

seq_documentation_block:
; Sequence 6, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405,496A
; FILING DATE: 16-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-405-496A-6

```

alignment_scores:

Quality: 85.00 Length: 196
Ratio: 0.825 Gaps: 9
Percent Similarity: 52.551 Percent Identity: 22.449

alignment_block:

US-09-323-427-3 x US-08-405-496A-6 ..

Align seg 1/1 to: US-08-405-496A-6 from: 1 to: 2710

```

529 TATCATTAATGACATGCGATTCTGAAACCGTTGACTTTCTGCGCGGT 578
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
188 TYTYTLYSSERGLINLEASNLYSPROTHVALPROTHIRLEASPSAII 204
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
579 TGCACATTCCTGCTTGTGATGATGTAACGGTGATGCTGGAATTC 628
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
204 eilelysserHISLeuvalSeriU.....T 213
629 TAAATGCTGATGATGCTGCTTGATTAATATTGCTAATAATTGAA 678
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
213 yrasnarGPSgluThrValleuGluSerTYrArgThrsnserleuArg 229
679 TATCCACAGATTTAATGCTGGCCAGAAAGCTACGTAACAAATATGC 728
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
230 .....Lys.IleA 232
729 GGATGATGACAGCTTTTCTATCAATGCCAGATCACTATTACCATTAAG 778
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
232 snBsrnHISglYIleASPIleArgAlaSnSerleuPhe..Thrglu 247
779 AACCAATAGCGAATGTTCGACCAATGTTCAAGACACACAGGATTC 828
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
248 GIngluLeuLeuasnIleTYrSerGIngluLeuLeuasnArgIleasnle 264
829 GGAAGCTGTTAAACAGGTGGTCCGACGCAAAACCTGTGACGTCGCA 878
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
264 uAlaAlaIleSerAspIleVal.....ArgleuLeuAlaIleuLysA 278
879 ACTTC.....GTTTACTCAAGAAAGATCTGCAGAACGGAGAAATATCA 922
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
278 snPheGlyIleValTYrLeuAspValAspMetleuProGlyIleHisSer 294
923 TTGATGTACGAACGTGATATCAACACC..... 949
295 AspLeuPheLysThrIleSerArgProSerSerIleGlyLeuAspArgTr 311
950 .....TTGAAATTAGCGATGATATCAAGCTTTCACAGTTGATT.... 988
311 pgluMetIleLysleuGluAlaIleMetLysTYrLysTYrIleAsnA 328
989 ..TACGTACCGGTGACATTC.....TGCAACATTAATGACACAC 1024
328 snTYrThrSerGluAsnPheAspLysLeuAspGlnGlnLeuLysAspAsn 344
1025 CTGTAAATACTGCTGCGATACAAATGCAATGTGCA 1060
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
345 PheLysLeuIleleuGluSerLysSerGluLysSer 356

```

seq_name: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:US-08-325-071-65

seq_documentation_block:

Sequence 65, Application US/08325071
Patent No. 5587311
GENERAL INFORMATION:
APPLICANT: COBON, Stewart Gary
APPLICANT: MOORE, Joanna Terry
APPLICANT: JOHNSON, Law Anthony York
APPLICANT: WILLADSEN, Peter
APPLICANT: KEMP, David Harold
APPLICANT: SRISKANTHA, Alagacane
APPLICANT: RIDING, George Alfred
APPLICANT: RAND, Keith No. 5587311man

TITLE OF INVENTION: DNA Encoding A Cell Membrane
TITLE OF INVENTION: Glycoprotein Of A Tick Gut
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/325,071
FILING DATE: 14-OCT-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,109
FILING DATE: 17-MAY-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/926,368
FILING DATE: 07-AUG-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU87/00401
FILING DATE: 06-JUL-1988

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU87/00401
FILING DATE: 27-NOV-1987

PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P14912
FILING DATE: 16-OCT-1987

PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P12570
FILING DATE: 19-JUN-1987

PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PH9196
FILING DATE: 27-NOV-1986

ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 60042/111 BIAU

TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300

TELEFAX: 202 672 5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 65:

SEQUENCE CHARACTERISTICS:

LENGTH: 620 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-325-071-65

alignment_scores:

Quality: 84.00 Length: 359
Ratio: 0.575 Gaps: 222
Percent Similarity: 40.669 Percent Identity: 21.448

alignment_block:

US-09-323-427-3 x US-08-325-071-65 ..

Align seg 1/1 to: US-08-325-071-65 from: 1 to: 620

```

138 TGCATTGGAAGACATGT.....TTATG 160
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
117 CysValProThrThrcysleuArgProAspLeuThrCysLysAspency 133
161 TGAAGAGCTTTATGATCAAGAAAGGTTGCCGTATATGATGAAGGTGACGT 210
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
133 sgluLysAsnleuLeuGlnIhrGasp.....SerA 143

```

```
211 CAAGTTGCCGGAATTCACCTTCATTCATTCGATGCGCGTAC 260
    |||:::|||||
143 rgycysgslnglytrpansThr.....AlasncysleuAlaIa 156
    ::::|
261 ACGATCTCGAATCCACGCTGCTATTTTGTAC..... 233
    ::::|
157 ProProAlaAspSer.....TyrCysSerProGlySerProLysG 170
294 .....ACAACTGTGTCTAT.....TTGC 312
    |||
170 yProAspGlyGlnCysLysAsnAlaCysArgThrLysGlnAlaGlyPheV 187
313 TTTCATCATTCATTTGTTTCCAAAGTTGATCGCATATCGAGTACAATG 362
    ||:::|||||
187 alCysLysHsGlyCys.ArgSerThrAspLysAlaTyrGlnCysThrCy 203
363 C.....TTTACATGGAAGCTGATAAACAAGTTAGTCACAGATG 403
    -:::|||||
203 sProSerGlySerThrValAlaGlnAspGlyIleThrCysLysSerIle 220
404 AGGTATCTGAAATCACACGCTTTCAAACTCGAATTTGC...CGATG 450
    ::::|
220 eTyrThrValSerCysThrValGlnGlnLysGlnThrCysArgProThr 236
451 CCAGTATGCCGTTATGAA.....ATTTGATGCTGACCAACCGG 491
    |||:::|
237 GluAspCysArgValGlnLysGlyThrValLeuGlnCysLysProThr 253
492 TCACCAAGTTCAATTGCTATCTATGCTCAGCCAGTTTATCATTAATGA 541
    ::|
253 ngIn.....HisLeuValGlyAsp.....L 260
542 CATCGCATTCCTGAACCGTTGATCTTCCGCGGTTGTCATCTCCTGC 591
    |||:::|
260 yscysIleSerAspCysValAspLysLysCys.....HisGlnGlu 273
592 TTTGTGATGATGATG..... 606
274 PheMetAspCysGlyValTyrMetAsnArgGlnSerCysTyrCysProTr 290
607 .....AACGTCATCTGCTGGAATTCGAATTCGATGATGAT 643
290 pLysSerArgLysProGlyProAsnVal.....AsnIleAsnGlnL 304
644 GTGCTCTTGATAATATTGCTTAAT.....AATTGGAA 678
    |||:::|
304 yslLeuLeuAsnGlnTyrTyrTyrThrValSerPheThrProAsnIleSer 320
679 TATCCAAACAGATTTAAATGCTGGCCAAAGACTCAGCTATACAATATGC 728
    :::|
321 PheAspSerAsp.....HisCysLysArgTyrGln 330
729 GGAATGCATCACACGCTTTTCATCATATCCGATCATATACCATTAAG 778
    |||
330 uAspArg.....ValLeuGlnAlaIleArgThrSerIleGlyLysGlnV 345
779 AACCAATATAGCGAATGTTCGACACAAATGTCAGAACACAGATTC 828
    ::|
345 alPheLysValGlnIleLeu.....AsnCysThrGlnAsp..... 356
829 GGAAGCTGTTAAACAGGTGGTCCGACGACAAACCTGCTGCAGCTGGCA 878
    ::|
357 .....IleLysAlaArgLeuIleAlaGlnLysProLeuSerLysTyrVa 371
879 ACTTCT..... 885
    |||
371 lleuArgLysLeuGlnAlaCysGlnHisProIleGlyGlnTrpCysMetM 388
886 .....TTACTCAAGAAAAGATCTGCA.....GAACCG 912
    |||
388 eTyrTrpLysLeuLeuIleLysLysAsnSerAlaThrGlnIleGlnGlu 404
913 GAGAAATATCATGTGATGATGACGAACCTGATATCAACACCTTGAAATTA 962
    |||
405 GluAsnLeuCysAspSerLeuLeuLysAsnGlnGlnAlaAlaTyrLysG 421
963 TGATATCAAGCTTTGCCAGTTGAT 987
    ::|
421 yGlnAsnLysCysValLysValAsp 429
```

11:11:15.1520

THIS PAGE BLANK (USPTO)


```

944 ACACCCCTGAATTTAGCGATGATATCAACCTTTGCCAGTTGATTTACGT 993
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
239 snthrlleuaspillemetgluLy.....AlaserProserAlaprogLu 233
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
994 CACCGTCACATTTCTGCACATATGACAAACCTGTAACTTGTCTGCAGT 1043
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
254 AlAlAlAlAlAlAlAlAlAlSerGlu.....GluSerValArgArgAlAlAl 268
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1044 ACAAAATGAGATCTGCATGTCACCATTTGGCTCTCAATGTTATGGGTT 1093
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
268 rserthnglylleserSerthProleGlyLeuProserPheleuGlym 285
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1094 TAAGCATTCGATTCGATTCGCCGCTATTTATACCATTTGTTAAATTT 1143
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
285 etArgThrlleValAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAl 301
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1144 GCMCA 1149
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
302 ArgPro 303
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: p1r2:SI7855

```

```

seq_documentation_block:
peptidylglycine monooxygenase (EC 1.14.17.3) - African clawed frog
N:Alternate names: peptidylhydroxyglycine N-C lyase
C:Species: Xenopus laevis (African clawed frog)
C>Date: 22-Nov-1993 #sequence_revision 03-Aug-1995 #text_change 18-Jun-1999
R:Iwasaki, Y.; Kawahara, T.; Shimoi, H.; Suzuki, K.; Ghisalba, O.; Kangawa, K.; Matsuo,
Eur. J. Biochem. 201, 551-559, 1991
A>Title: Purification and cDNA cloning of Xenopus laevis skin peptidylhydroxyglycine N-C
A:Reference number: SI7855; MUID:92037609
A:Accession: SI7855
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-935 <IWA>
A:Cross-references: EMBL:X62771; NID:g64530; PIDN:CAA4615.1; PID:g64531
C:Superfamily: peptidylglycine monooxygenase II; peptidylglycine monooxygenase I homolog
C:Keywords: oxidoreductase
F:131-342/Domain: peptidylglycine monooxygenase I homolog <PCM>

```

```

alignment_scores:
  quality: 100.00      length: 391
  ratio: 0.541         gaps: 18
  percent similarity: 47.315      percent identity: 19.437

```

```

alignment_block:
US-09-323-427-3 x SI7855 ..

```

```

Align seg 1/1 to: SI7855 from: 1 to: 935

```

```

97 TGTGGACCACTTCATTAACATCAATTTTATACAGCTAATGCTTGA 146
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
526 CysGlnProthraspAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAl 542
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
147 AGAACAATGTTATGTAAGAGCTTTATGATCAAGAAGTTGCCGTAATG 196
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
542 lAlAlAspGlyTyr..... 546
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
197 ATGAAGGTGAGCTCAAGTTGCCGAATTTCACTTCATTGATTCATGC 246
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
547 .....Cys 547
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
247 AATGTTCCGCTACAGATCTCTGAATCCACGCGTATTTTGTGA..... 291
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
548 Asp...SerArglleMetGlnPheSerProAsnGlyMetPheIleMetG1 563
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
291 ..... 291
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
563 nTPbGlyGluGlnThrSerSerAsnValProArgProGlyGlnPheArg1 580
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
292 .....ACAACTGTTGTGTCATTTGTTGATTCATTCATTATTTGTTACC 333
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

580 leProHisSerleuThrMetValProAspGlnGlyGlnLeuCysValAlAl 596
      ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
334 AAGTTCATGTCGATATATGACATCAATGCTTTTACATGGAACCTGATA 383
      ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
597 AspArgGlnAsnGly...ArglleGlnCysPheHisAlAlAlAlAlAlAl 612
      ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
384 AACAGTTAGTCACAGATTCAGGTATCTGAAATC...ACAACTGCTTTTC 430
      ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
612 nPheValLys...GlnlleLysHisGlnGlnPheGlyArgGluValAlPhea 628
      ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
431 AAACGCAATTTGCCGATGCCAGTATGCCGTTATGAAATTTTGATGGT 480
      ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
628 lAlAlSerTyrAlAlPro.....Gly 634
      ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
481 GGACCAACCGGTCACACGATTCATTTGCTATGATTCGTCACGCACTTGA 530
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
635 Gly.....ValleuTyrAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAl 646
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
531 TCATTAATGACATGCGATTCGAAACCGTTGATTCCTGCGCGGTTG 580
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
646 r.....GlyTyrSerAlAlPro 652
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
581 TCCATTCCTGCTTTGTCGATGATGTAACGCGTATCTGGAATTCGA 630
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
652 alGlnGlyPheMetleuAsnPheSerAsnGlyLysP..... 663
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
631 AATGCTGATGATGCTGCTTGAATATTTG.....CTAATATTTT 674
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
664 .....lleuAspThrPheIleProAlAlArgLysAsnPh 675
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
675 GGAATATCCACAGATTTAATGGCTGGCCAAAGACGCTATACAAAT 724
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
675 eAspMetProHisAspIleAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAl 692
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
725 ATGCGATGATGATGATGCTTTTATCAATGCCAGATCACT..... 765
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
692 lAspAlAlHisAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAl 708
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
766 ...ATTACCTTTAAGAACCAAT.....ACGGAATGCTTTGAGCC 803
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
709 HisArgSerValLysLysAlAlGlylleGluValGlnlleThrGluTh 725
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
804 ACAATGTTCAAGACCAACAGATTCGAGCTGTTAAACAGTGGTCCG 853
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
725 rGluIlePheGlnThrHislleArgSerArgProLysThrAsnGlnSerV 742
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
854 CAGCAAAACCTGCTGCAGCTGCGCAACTTCGTTTACTCAAGAAAGATCT 903
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
742 alGluLysGlnThrGlnGlnLysGlnGlnLys.....GlnLysAsnSer 756
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
904 GCA.....GAACCGAGAAATATCATTT...GATGTACGAAC 935
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
757 lAlGlylValSerThrGlnGlnLysGlnAsnValAlAlGlnlleAsnAl 773
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
936 TGAATATCAACACCTTGAA.....ATTACGATGATTAATCAACGCTT 976
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
773 aGlyValAlProThrGlnGlnLysGlnAsnValAlAlGlnlleGlnSerAlaG 790
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
977 TGCAGTTGATTCGTCACCGCTGCACCTTCCTGCAACATATGACAAACCT 1026
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
790 lValSerThrGlnGlnLysGlnSerValAlAlGlnlleGlnSerAlaGly 806
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
1027 GTAATACTTCCTGCAGTACAAATGGAATGATGATGTCACCATTTGGCTT 1076
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
807 ValSerThrGlnGlnLysGlnSerValAlAlGlnlleGlnSerAlaGly 823
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
1077 CTCATGCTTATGAGGTTTAAAGATTCATTCATTTGCTGCGCGTATTA 1123
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
823 lSerPheValleuIlelleThrleuLeuIlelleProIleAlAlValleuI 840
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
1124 TTACCATTTGCTTTAAATTCGT 1146
      ||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:

```


844 GGTGGTGGCCG.....AGCAAAACCTGCTGC... 869

```

E:/64-78//Domain: transmembrane #status predicted <TMN>
E:/65-662-743//Binding site: carbohydrate (Asn) #status predicted

```

[illegible]

584 leProHisSerLeuThrMetIleSeraspGlnGlnLeuGlnCysValAla 600

```

334 AAAGTTGATCGTCATATCGAGTACAAATGCTTTTACATGGAAGCTGATA 383
      :::::::::: ||:::|||||::: ::::: ::
601 Asparaginyl... ArgileGlnCysPheHisAlaLysThrGly 616
384 AACGTTAGTGCACAGATTGAGTATCTGAATC..ACAAGTCCTTTTC 430
      ||::: ||::: ||::: ||::: ||::: ||:::
616 uPheValLys... GlnLeuLysHisGlnIlePheGlyArgGluValPhe 632
431 AAAGTCAAAATTTGCCGATGCGACGATGCGGTATGAAATTTTGATGAT 480
      ::::: ||::: ||::: ||::: ||::: ||:::
632 LysAlaSerThrAlaPro.....Gly 638
481 GGACCAACCGGTCAACGATTCATTTGCTATCATGTCAGCAGCTTTTA 530
      ||::: ||::: ||::: ||::: ||::: ||:::
639 Gly.....ValLeuThrAlaValAlaSnGlyLysProGly 650
531 TCATTAATGCACATGCATTTCTGAACCGTTGATCTTTCGCGCGGTG 580
      ||::: ||::: ||::: ||::: ||::: ||:::
650 rGly.....AspSerThrProValGlnGlyPhe..... 659
581 TCATTCCTGCTTGTGATGATGATGATGATGATGATGATGATGATGAT 630
      ::::: ||::: ||::: ||::: ||::: ||:::
660 .....MetLeuAsnPheSerAsnGlyAsp..... 667
631 AATGCTGATGATGATGCTCTTGAATAATATTG.....CTAATAATT 674
      ||::: ||::: ||::: ||::: ||::: ||:::
668 .....IleLeuAspThrPheIleProAlaArgLysAsnPh 679
675 GGAATATCCACAGATTATATGCTGGCCAAAGATCAGCTATATC 720
      ||::: ||::: ||::: ||::: ||::: ||:::
679 eGluMetProHisAspIleAlaIleGlyAspAspGlyThrValTyr 694

seq_name: p1r2:T15881

```

```

seq_documentation_block:
  hypothetical protein D1044.3 - Caenorhabditis elegans
  C:Species: Caenorhabditis elegans
  C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
  C:Accession: T15881
  R:Pauley, A.
  submitted to the EMBL Data Library, June 1994
  A:Description: The sequence of C. elegans cosmid D1044.
  A:Reference number: Z18423
  A:Accession: T15881
  A:Status: preliminary; translated from GB/EMBL/DBJ
  A:Molecule type: DNA
  A:Residues: 1-1895 <PAU>
  A:Cross-references: EMBL:U00065; NID:9495681; PID:9495684; PIDN:AAA50735.1; CESP:D1044.3
  A:Experimental source: strain Bristol N2
  C:Genetics:
  A:Gene: CESP:D1044.3
  A:Introns: 35/2; 63/2; 150/1; 191/3; 207/2; 405/3; 551/3; 629/1; 764/1; 933/1; 959/2; 10

```

```

alignment_scores:
  Quality: 94.00      Length: 209
  Ratio: 0.940      Gaps: 8
  Percent Similarity: 47.847      Percent Identity: 21.053

alignment_block:
  US-09-323-427-3 x T15881 ..
  Align seg 1/1 to: T15881 from: 1 to: 1895

246 CAATGTTGGCCGACACGATCTCGAATCCACGCGATTATTTGTACAA 295
      ||||| ::::: ||::: |||||
1149 GlnCysIleAspAsnSerValCysMetAlaGlnMetCysThrCysAsnAs 1165
296 CAAC.....TGTTGTATTTGGTTTCATCATTAATT 327
      ||| |::: ||::: ||::: ||::: ||::: ||:::
1165 nAsnTyrAlaGluValTyrGlyTyrCysValProIleHisSerIleC 1182

```

```

328 GTTACCAAGTTGATCGTCATATCGAGTACAAATGCTTTTACATGGAAGC 377
      || ||::: ||::: ||::: ||::: ||::: ||:::
1182 yGlnGlnIleThrGlnThrLeuVal..... 1189
378 TGATAAACAAGTTAGTGCACAGATTGAGTATCTGAATCACAACTCCTT 427
      ||::: ||::: ||::: ||::: ||::: ||:::
1190 .....AsnAsnGlnCysValLeuLeuSerIleValGlyGlnThrCys 1204
428 TTCAACTCAAAATTTGCCGATGCGACGATGCGGTATGAAATTTTGAT 477
      ::::: ||::: ||::: ||::: ||::: ||:::
1204 eAlaAsnGlnGlnCysValGlyAlaAlaMetCys.....AsnSerGly 1219
478 GGTGACCAACCGGTCAACGATTCATTTGCTATCATGTCAGCAGCT 527
      ::::: ||::: ||::: ||::: ||::: ||:::
1219 hrcysGlnCysThrAsnGlyAlaThrAlaMetTyrGlyTyrCysIleSer 1235
528 TTATCATTAATGACATGCATTTCTGAACCGTTGATCTTCTGCGCGG 577
      ||::: ||::: ||::: ||::: ||::: ||:::
1236 SerSer.SerSerSerCysAsnSerAsnGlnVal.....Ser 1248
578 TTGTCATTCCTGCTTGTGATGATGATGATGATGATGATGATGATGAT 627
      ::::: ||::: ||::: ||::: ||::: ||:::
1248 IeAsnGlyMetCysTyr.....AsnThrValGlnVal 1258
628 CTAATGCTGATGATGATGCTCTTGAATAATATTGCTAATAATTTGGA 677
      ::::: ||::: ||::: ||::: ||::: ||:::
1259 .....GlyGlySerCysSerPheSerGlnGlnCysLeuAsnAlaVal 1273
678 ATATCCACAGATTATATGCTGGCCAAAGATCAGCTATACAAATATG 727
      ||::: ||::: ||::: ||::: ||::: ||:::
1273 LysThrAsnAsnIleCysValSerThrPheCysSerValSerCysSer 1290
728 CGGATCGATCAGACGCTTTCTATCATCCAGATCATGATTAACATTAA 777
      ::::: ||::: ||::: ||::: ||::: ||:::
1290 hIrsnGlnValCysIleSerAsnGlnCysTyrAsnTyrValSerIleGly 1306
778 GAA.....CCAATAGCGAATGTGCTGC 800
      ::::: ||::: ||::: ||::: ||::: ||:::
1307 SerGlnCysValGlySerGlnGlnCysLeuSerAsnSerGlnCysIle 1323
801 ACCCAATGTTCAAGCCACCAAGCA 825
      ||::: ||::: ||::: ||::: ||::: ||:::
1323 rSerIleCysGlnCysProGlnGly 1331

seq_name: p1r2:C70126

```

```

seq_documentation_block:
  DNA mismatch repair protein (mutL) homolog - Lyme disease spirochete
  C:Species: Borrelia burgdorferi (Lyme disease spirochete)
  C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 05-Jun-1998
  C:Accession: C70126
  R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
  son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Yu
  ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
  Nature 390, 580-586, 1997
  A:Authors: Smith, H.O.; Venter, J.C.
  A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
  A:Reference number: A70100; MUID:98065943
  A:Accession: C70126
  A:Status: preliminary; nucleic acid sequence not shown; translation not shown
  A:Molecule type: DNA
  A:Residues: 1-610 <KLE>
  A:Cross-references: GB:AE001131; GB:AE000783; NID:92688098; PID:92688099; TIGR:BB0211
  A:Experimental source: strain B31

```

```

alignment_scores:
  Quality: 93.50      Length: 389
  Ratio: 0.519      Gaps: 23
  Percent Similarity: 46.272      Percent Identity: 22.879

alignment_block:
  US-09-323-427-3 x C70126 ..

```

[illegible]

seq_name: pir2:T11616

seq_documentation_block:

C:Species: Schizosaccharomyces pombe

C;Accession: T11616

submitted to the EMBL Data Library, October 1995

A;Accession: T11616

A: Molecule type: DNA

A/Cross-references: EMBL:Z54328; NID:g1009451; PID:g1009456

C;Genetics:

A;Note: SPAC22G7.06c

C;Keywords: ligase

alignment_scores:

Quality:	92.00	Length:	365
Ratio:	0.544	Gaps:	174
Percent Similarity:	46.301	Percent Identity:	20.274

alignment_block:

US-09-323-427-3 X T11616

Align seg 1/1 to: T11616 from: 1 to: 2244

```

124 TTTAATACAGCTGATTCGACGACGACATGTTATCTGAAAGCTCTTA 173
    : : : : : | | | | | | | | : : : : :
999 TTTThTThTtTyrAsnAlaValGluHHisAspLeHisphe..... 1011111
174 TGTATCAAGAAGTGTGCCGTAAATGATGCAAGT.....G 205
    : : : : : | | | | | | | | : : : : : |
1012 .....AsnAspTyrGlyValMetValLeuLysG 10222
206 GACGTCACACTGGCCGAATTTCACTTCCTCCATTGATTACGACATGTTGCG 255
    : : : : : | | | | | : : : : : | | | | |
1022 lValTtYrAlglIegIysSerValGluIbheAspTyrCysAlaValAla 10388
256 CGTACGACATCTGTGAATCCACGCTGGATTTTGTATACACACACTGTGTT 305
    : : : : : | | | | | : : : : : | | | | |
1039 AlAlaValTThrLeuAlaTyrGAspArgGlyVal.....LysTThrIle 10522

```



```

354 rPThrIaAsnCysHisLysCysThrCysThrAspAlaGluThrValAsp 370
565 .....ACTTCGCGGGGTTGCATTCCTGCTTTCATGATGG 605
371 CysLysLeuLysGluCysProSerProProThrCysLysProGluGlu.. 386
606 TAACGGTGATGACTGTGAATTTCTAAATGCTGATGATGCTGCTTGATA 655
387 .....ArgLeuValLysPheLysAspAsnAspThrCysCysGluLeu 401
656 AATAT.....TTGCTAAATTAATTTGGAATATCCACA 687
401 LetyrCysGluProArgThrCysLeuPheAsnAsnAspTyr..... 415
688 GATTATATGGCTGGCCAGAGCTCACTATACAAATATGCGATGATC 737
415 ..... 415
738 ACAGCTTTCTATCATATGCCAGATCATATTACATTAAAGACCAATA 787
416 .....GluValGlyAlaSerPheAlaAspProLysA 426
788 GCCAATGTGTTCCACCAATGTTCCAGAACACAGATTCGAGCTGT 837
426 snProCysIleSerTyrSerCysHisAsn...ThrGlyPheValAlaVal 441
838 AA.....AACAGTGGTGGCC..... 854
442 ValGlnAspCysProLysGlnThrTrpCysLagLagLysAspArgValTyr 458
855 .....ACGAACCTGCTGAGCTGCGCACTGCTTACTACAGAAAGA 900
458 rasPserThrLysCysCys.....TyrThrCysLysProt 470
901 TCTGCAGAACGGAGATATATCATGATGATGACAC 935
470 yrcysargserServalAsnValThrValAsn 481

seq_name: p1r2:S72278

```

```

seq_documentation_block:
ATP-dependent Clp proteinase (EC 3.4.21.-) homolog - Plasmodium falciparum plastid
C:Species: Plasmodium falciparum
C:Date: 14-Apr-1998 #sequence_revision 24-Apr-1998 #text_change 07-May-1999
C:Accession: S72278; S78483
R:Wilson, R.J.M.; Denny, P.W.; Preiser, P.R.; Rangachari, K.; Roberts, K.; Roy, A.; Whyt, J. Mol. Biol. 261, 155-172, 1996
A:Title: Complete gene map of the plastid-like DNA of the malaria parasite Plasmodium fa
A:Reference number: S72277; MUID:96346169
A:Accession: S72278
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-765 <WIL>
A:Cross-references: EMBL:X95276
R:Wilson, R.J.M.
A:Reference number: S78483
A:Accession: S78483
A:Molecule type: DNA
A:Residues: 1-95, 'N', '96-765 <WIM>
A:Cross-references: EMBL:X95276; NID:g1171591; PID:g220210; PID:g1171612
C:Genetics:
A:Gene: clpC
A:Genome: plastid
A>Note: this apparently degenerate plastid is referred to as the apicoplast
C:Keywords: hydrolase; plastid; serine proteinase

```

```

alignment_scores:
Quality: 90.00 Length: 225
Ratio: 0.720 Gaps: 20
Percent Similarity: 55.556 Percent Identity: 27.111

```

```

alignment_block:
US-09-323-427-3 x S72278
Align seg 1/1 to: S72278 from: 1 to: 765

```

```

314 TTCATCCATTATTTGTACCAAGTTGATC.....GTGCATATC.. 352
282 TyrLeuPheLeuLeuLeuAsnLysLeuTyrGlyTyrAsnIleHisIle 298
353 .....CAGTACATGCT..TTTACATGGAACGTGA 380
298 eileValThrAsnLysLysGluTyrAsnThrTyrPheLysTyrAsnIle 315
381 TAAACAG.....TTAGTGACAGATTGAGTATCG 412
315 LelysAspSerTyrPheTyrLysIleArgIleLysAspSerIleLeu 331
413 AAATCAACAAGCTGTTTCAACATCAATGTGCCGATGCCGATGCCGT 462
332 GlnThrPheLeuIleLeuLysAsnAsnIle.....TyrLysTyr 344
463 TATGAATTTTGGATGGTGACCAACGGTCAACAG.....TTCA 503
344 rIleAsnTyrTyrLysIleAsn.....IleAsnAsnTyrIleIleTyr 359
504 ATTTGCTATCATTTGTCAGCCAGTTATCATTAATGACATGCGATTG 553
359 IuLeuIleAsnLysSerLysLysTyrIle..... 368
554 AAACGTTGATCTTCTGCGCGGTGTCATTCCTGCTTTCGATGAT 603
369 LysProLeuIleLeu.....ProThrThrProLeu.. 378
604 GGTAAAGGTGATCTGTGAATTCGAATGCTGATGATGCTGCTTGA 653
379 .....IleLeuLeu.GluAsnSer.....CysSer.As 387
654 TAAATATTTGCTAAATAT..TTGGAATATCCACAGATTATATGGCTG 700
387 nLysTyrLeuLeuAsnAsnLysIleSerTyr.SerAsnPheAsnTyrLeu 403
701 GCCAAGAGCTCACTATACAAATATGCGATGATGACACGTTTCTAT 750
404 .....PheThrTyrAsnAsn...AsnIleIleTyrAsnAsnLysAs 416
751 CAATGCCAGATCAGTATATCCATTAAAGAACCAATAGCGATGTGTCG 800
416 nasAsnLeuThrIleGluAspIleLysAsnSerIleSerAsnTyrLeu. 432
801 ACCCAATGTTCCAGAAC.....CACAGATTCGAGCTGTTAAAA 841
433 .....AsnIleSerLysThrIleLeuPheLysAspAsnLysLeuThrLys 447
842 CAGGTGTCGCCGAGCAAAACCTGTCGACCTGCGCAACTGCTTATAC 891
448 .....LeuAsnLeuThrLysLeuGlnAsnTyrLeuTyrAs 459
892 AAGAAAGATCTGCAGAAC 910
459 nHisIleTyrGlyGlnAsn 465

seq_name: p1r2:S48944

```

```

seq_documentation_block:
hypothetical protein YHR102w - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 13-Jan-1995 #sequence_revision 10-Feb-1995 #text_change 24-Sep-1999
C:Accession: S48944
R:Latterille, P.
A:Submitted to the EMBL Data Library, May 1994
A:Description: The sequence of S. cerevisiae cosmid 8263.
A:Reference number: S46691

```

A:Accession: S48944
 A:Molecule type: DNA
 A:Residues: 1-1080 <LAT>
 A:Cross-references: EMBL:U00059; NID:g529116; PION:AB6860.1; PID:g529127; MIPS:YHR102w
 C:Gene: SCD:NRK1
 A:Cross-references: SCD:S001144; MIPS:YHR102w
 A:Map position: 8R
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
 C:Keywords: ATP; serine/threonine-specific protein kinase
 F:21-276/Domain: protein kinase homology <KIN>
 F:29-37/Region: protein kinase ATP-binding motif

alignment_scores:
 Quality: 90.00 Length: 209
 Ratio: 0.833 Gaps: 6
 Percent Similarity: 51.675 Percent Identity: 21.531

alignment_block:
 US-09-323-427-3 x S48944 ..

Align seg 1/1 to: S48944 from: 1 to: 1080

```

430 CAACCTCAAAATGTCGCCGATGCCAGTATGCCGTTATGAAATTTGGATGG 479
      :::::|||||
      :|||:::|||||
200 LysValAspIleThrSerLeuGlyIleThrThyGluIleAlaThrG1 216
      :|||:::|||||
480 TGGACCA.....ACCGGTCAACCGTTCATTTCTTA 511
      :|||:::|||||
216 yasnProPrioTyrCysAspValGluAlaLeuArgAlaMetGluLeuIle1 233
      :|||:::|||||
512 TCATTTGGTCGACGATTATCATTAATGACATGCGATTCGTAACCGTT 561
      :|||:::|||||
233 IeYssSerIysProPrioArgLeuGluAspArgSerIyrSerThSerLeu 249
      :|||:::|||||
562 GATACTTTTCGCGGGTGTTCATTCCTGCTTGTGATGATGTTAAGG 611
      :|||:::|||||
250 LysGluIleAlaLeu.....CysLeuAspGluAspProLys.. 262
      :|||:::|||||
612 TGAATACGTGGAATTTCTAAATGCTGATGATGCTCTTGATTAATAT 661
      :|||:::|||||
263 .....GluArgLeuSerAlaAspAspLeuLeuLysSerIysPheI 276
      :|||:::|||||
662 TGCTAAATATTTGGAATATCAACAGATTAATGCTGCGCAAGAGCT 711
      :|||:::|||||
276 IeArgAlaHisLysAlaThrProThrSerIleLeuLysGluLeuIleSer 292
      :|||:::|||||
712 CACGTATACAAATATGCGGATCGATCAGACGTTTCTATCAATGCCAGAT 761
      :|||:::|||||
293 ArgTyrIleLeuIlePheArgAspLysAsnLysAsnLysIleGluG1 309
      :|||:::|||||
762 CAGATATACATTAAGAAGAACCAATAGCGATGCTCGACCAACAGATT 811
      :|||:::|||||
812 CAGAACCAAGAGATTGCGAGCTGTTAAACAGGTGCGCGCAACAAA 861
      :|||:::|||||
326 erGln.....AsnLysGlyGlyAspGluAlaGln 335
      :|||:::|||||
862 CCGTCTGACGTCGCGCAACTGCTTTACTCAGAAAAAGATCTGCAGAAC 911
      :|||:::|||||
336 LysSerIleAlaSerAsnAspAsnGluIleLysArgValAsn..... 349
      :|||:::|||||
912 GGAGAAATATGATGATAGCAACTGATATCAACACCTTGAATTTAGCG 961
      :|||:::|||||
350 .GluGlyAspValGluMetLysTrpAspPheAspSerLeuSerSerSer 366
      :|||:::|||||
962 AT.....GATATCAGCTTTCGACAGTTGATTTA 990
      :|||:::|||||
366 spIyrIleIleGluAsnAsnIleAsnLeuAspAlaLeuAlaGluAspAsn 382
      :|||:::|||||
991 CGTACCGTGCACCTTTCGAAACATTAAT 1017
      :|||:::|||||

```

```

383 AsnGluTrpAlaThrAlaGlnHisAsp 391
      :|||:::|||||
seq_name: p1r2:C71618

```

seq_documentation_block:
 hypothetical protein PFB0315w - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 07-May-1999
 C:Accession: C71618
 R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.;
 .; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.
 A:title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference number: A71600; MUID:99021743
 A:Accession: C71618
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1712 <GAR>
 A:Cross-references: GB:AE001386; GB:AE001362; NID:g3845148; PID:g3845151; TIGR:PFB031
 A:Experimental source: clone 3D7
 C:Gene: PFB0315w

alignment_scores:
 Quality: 90.00 Length: 341
 Ratio: 0.584 Gaps: 12
 Percent Similarity: 45.161 Percent Identity: 19.941

alignment_block:
 US-09-323-427-3 x C71618 ..

Align seg 1/1 to: C71618 from: 1 to: 1712

```

173 ATGATCAGAAGCTGCCGTAATGATGACGTGAGCTCAAGTTGCCGA 222
      :|||:::|||||
      :|||:::|||||
1102 MetLysLysLysLysLysThrGlnGluTyrValAspIleGlu.....Th 1116
      :|||:::|||||
223 ATTTCACCTTCATTTGATTCATGCAATGTTCCGCGTACAGACTCGAA 272
      :|||:::|||||
1116 rValTyrGluTyrValIleGluLysTyrLysArgValHisValLeuTyrL 1133
      :|||:::|||||
273 TC...CACGTGATATTTTGTAAACACACACTGTTGTCATTTGTCATC 319
      :|||:::|||||
1133 euGlyArgLeuGluGlnIleValGluLysLeuPheLysLysTyrIleLeu 1149
      :|||:::|||||
320 CATTAATTTTACCAAGTTGATCGTCATATGAGTACATGCTTTTAC 369
      :|||:::|||||
1150 LysTyrSerPheHisLysLeuArgIlePheTyrGluTyrLysIleGluMe 1166
      :|||:::|||||
370 ATGGAAGCTGATTAACACGATTAGTGCACAGATGAGTATCTGAAATCAC 419
      :|||:::|||||
1166 tGluLysLeuLysLysAsnTyrIleHisCysIleTyrAspIleSerSpl 1183
      :|||:::|||||
420 AACTGCTTTTCAAACTCAAAATGTCGCGATGCGAGTATGCGTTATGAA 469
      :|||:::|||||
1183 yLeuGluPheLeuIleLysLysLysMetGlnHisTyr..... 1195
      :|||:::|||||
470 TTTTGGATGCTGACACACCGGTCAACCCAGTTCAATTTGCTATCATTTG 519
      :|||:::|||||
1196 .....PheAsnHisIleIleIleAsnSerTyrG1 1205
      :|||:::|||||
520 CACGCACTTATC.....ATAATGACATGCGATTCGTAACCCGT 560
      :|||:::|||||
1205 uSerSerPheIleAsnTyrGlnIleLysTrpAsnAspMetLeuTyrAsnL 1222
      :|||:::|||||
561 TGATACTTTTCGCGGCTGTCATTCCT...GTTTGTGATGATAGTGA 607
      :|||:::|||||
1222 euLeuLeuLysGluLysSerAlaTyrGlnAsnHisLeuGlyLysAsnTyr 1238
      :|||:::|||||
608 ACGGTGATTAAGTGAATTTCTAAATGCTGATG.....GATGCTCTTT 651
      :|||:::|||||

```

```

1239 IleuIleuTyrIlySValIleuLeuSerMetTyrIlySlySMeAl 1255
652 GATAAATATTGCTGAATATTTGGAAATATCCAAAGATTTAAATGCGTGG 701
1255 alleTyrPheArgSerPheValIlyrSnaSnIleIySValSerIlySlyL 1272
702 CCAAGAGCTCAGGTATACAAATATGCGATGCATCAC..... 739
1272 ysaSnAlaPheAlaTyrThrLeuThrArgValaSnSerIleuValLeu 1288
739 ..... 739
1289 TyrGluArgArgIleuYSerPheIlePheSerIlySleuYSpheAsnIy 1305
740 .....ACGTTTCATCAATGCCGAT 761
1305 rAspAsnValSerTyrPheCysPheThrMetTyrIlySIIeTyrLeuArgA 1322
762 CAGTATT.....CCATTAAAGAACCAATAGCG 790
1322 rglLeuPheGlyTyrLeuArgIleArgAspAsnArgIleAsnIleIys 1338
791 AATGTGTGCACCAATATTT..... 811
1339 AsnValIleGluIlySAsnValTyrArgLeuValIlySleuIleSerIlySII 1355
812 ....CAGAACCAAGAGATTCGAGCTGTAAACAGAGTGTGCGCGCAGC 857
1355 eSerAspAsnHisIlyStryrSnaIlePheIlyS..... 1366
858 AAAACCTGCTGCAGCTCGCACTTCGTTTACTCAAGAAAGATCTGCAG 907
1367 .....LeuGlnIlyStryrValIlyrGluGlnAsnGluIlySlys 1378
908 AACCGAGATATCATTTGATGTACGAACATATCAACACCTTGAATT 957
1379 AsnIlySMeIle..... 1382
958 AGCGAGATATATCAAGCTTTC...CAGTTGATTACGTCACCGTGCACT 1004
1383 .....CysAspAsnLeuIleTyrAlaAsnAsnGluL 1393
1005 TCTGCAACATATGACACACCTG 1027
1393 euCysAsnAsnLeuAspIlySIIe 1400

seq_name: p1r2:A41519

seq_documentation_block:
posterior-group protein tudor - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 24-Sep-1999
R:Golumbski, G.S., Bardsley, A.: Tax, F.: Boswell, R.E.
Genes Dev. 5, 2060-2070, 1991
A>Title: tudor, a posterior-group gene of Drosophila melanogaster, encodes a novel prote
A:Reference number: A41519; MUID:92038995
A:Accession: A41519
A:Molecule type: mRNA
A:Residues: 1-2515 <GOL>
A:Cross-references: GB:X62420; MID:98753; PIDN:CAA44286.1; PID:98754
A:Gene: tud
A:Genetics:
A:Cross-references: FlyBase:FBgn0003891
C:Superfamily: posterior-group protein tudor

alignment_scores:
Quality: 90.00 Length: 367
Ratio: 0.552 Gaps: 16
Percent Similarity: 44.414 Percent Identity: 19.074

alignment_block:

```

```

US-09-323-427-3 x A41519 ..
Align seg 1/1 to: A41519 from: 1 to: 2515

46 TCTTATTTCATTCGCGTGTCGAATGCTGCGAAGGTGAGCCGAATTTGA 95
111 ..... 111
1904 SerLeuGlnLeuPro...AspAlaTyrIleSerTrpSerProGluAlaG 1919
96 ATGTGGACCAACTTCGAATACA.....ATCAATTTTAATACAC 133
111 ..... 111
1919 uAlaIlySPhelAlaGluLeuThrGlyGluLeuValPheThrThrG 1936
134 GTAATGCATTCGAAGA.....CATGTTATGTGAAGGCTTTATGAT 177
111 ..... 111
1936 InLeuLeuIlySProGlyGlnAspHisValThrIleAspLeuLeuAsp 1952
178 CAAGAAAGTTGCCGTAATGATGAAGGTGACGTCGAAGTTGCCGAATTTC 227
111 ..... 111
1953 .....GlyGluAsnIleIleAspArgLe 1960
228 ACCTTCATTTGATTTCATGCAATGTTGGCGGTACAGCATCTGAATCCAC 277
111 ..... 111
1960 uLeuProLeu.....CysGlnArgIlySglnProIlySglnAlaSerIlySg 1975
278 GTGGTATTTTGTACACACACTGTGTCAATTCGTTTCATTCATTATT 327
111 ..... 111
1975 IuSerLeuAlaValThrThrIlySAlaIleIle..... 1985
328 GTTACCAAAATTTGATGTCGATATCGAGTACATGCTTTACATGGAAGC 377
111 ..... 111
1986 ....ThriHisValGlnAsnThrSerArgIleTyrLeuGlnPheSerGluIy 2001
378 TGAT..... 381
2001 sAspSerLeuMetAspIleIleCysGlnIlySleuAsnGlySerIlySleuG 2018
382 ....AAACAGTTAGTGCACAGATTGAG..... 405
2018 InProIlySThrGluIlySAlaIleValAspIlySmetCysValIleGlnPhe 2034
406 ..... 425
2035 AlaAspAspLeuGlnPheTyrArgSerArgIleLeuGlnValIleuGlnAs 2051
426 TTTTAAACACTCAAAATGTGCCGATGCCGATATGCCGTTATGAATTTGG 475
2051 pAspGlnTyrIlySValIle.....LeuIleA 2060
476 ATGTGTGACCAACC.....GGTCAACCAAGTTCAA 504
111 ..... 111
2060 sPtyrGlyAsnThrThrValIleValAspIlySleuTyrGluLeuProGlnGlu 2076
505 TTTGCTATTCATTTGTCAGCCAGTTTATCATAAATGGACATGCCATTCGCA 554
111 ..... 111
2077 PheThrLeuIle...LysProValAlaGlu.....IleCysSerMetG 2090
555 A.....ACCGTGAATACCTTCT 571
111 ..... 111
2090 uProSerAlaIlePheGlnIlySAsnIlySAlaLeuThrIleThrPhea 2107
572 GCGCGGTGTGCATTCCTGCTTGTGCGATGATGTAACGGTGATCTGTG 621
111 ..... 111
2107 sPAlaLeuLeuAspSerCys..... 2113
622 GAATTTCTAATGCTGATGATGTGCTCTTGATTAATATTGCTAAATAA 671
111 ..... 111
2114 .....LysGlyValAlaIleValAlaGluPheValAsnIlySse 2125
672 TTTGAATATCCAAAGATTTAATGGCTGCGCAAGAACTCAGATATACA 721
111 ..... 111
2125 rAlaSerProProValValArgLeuThrThrIlySAspIlySArgSerLeuL 2142
722 AATATGCGATGCATCACAGCTTTTCTATCAATGCCAGATCAGATTATACC 771

```


804 ACAATTCTTCACAAACCAAGAGTTCTGGAGCCTTTAAACAGAGTGC... 850
 |||||
 848 spsnllleaspllePheValGlyIleValIleuIleThrIleGluIleGlu 864
 851 ..CCGAGCAAAAACCTCTCGAGCTGGCGCAACTTGTTACTCAAGAA 897
 |||||
 865 GluProGluGluSerPheLeuGluIleuLeuAsn.....ArgI 877
 898 AGATCTGCAGAACCGGCAAAATTCATTGATGTACGACTGATATCAACAC 947
 :|||
 877 uasPleuAspIysArg..... 882
 948 CCTTGAATTTAGCGATGTAATCAAG 973
 :|||
 883 ..ValGlyHisAlaMetIleMetIys 890

THIS PAGE BLANK (USPTO)

alignment_scores: Length: 416
 Ratio: 4.228 Gaps: 7
 Percent Similarity: 80.769 Percent Identity: 67.067

alignment_block:

US-09-323-427-3 x CUI1_CAEEL ..

Align seg 1/1 to: CUI1_CAEEL from: 1 to: 424

```

25 TGTACTACACTTATTCATTCCTTATTCG...ATCCGGTTGCACATCG 71
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
   8 CysLeuAlaLeuValLeuSerAlaSerAlaLeuProValAspAsnAs 24
72 TGTGAGAGGTGAGCCAGAAATTCATTCATTCGACACACTTCATTAACAATCA 121
   : : : : : : : : : : : : : : : : : : : : : : : : : :
   24 nValGluGluProGluValGluCysGlyProAsnSerIleThrValA 41
122 ATTTTAATACAGTAAATGCATTCGACAGACATCTTTATGTGAAAGTCTT 171
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
   41 snPheAsnThrArgAsnProPheGluGluHisValTyrValLysGlyLeu 57
172 TATATGACAGAGGTTCGCGTAATGATGAGAGGTGAGGTCAAGTTGCCGG 221
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
   58 TyrAspGlnAlaGlyCysArgSerAspGluGlyTyrGlnValAlaG 74
222 AATTTCACCTTCATTCATTCATTCATTCGCGGTACAGACATCTCGA 271
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
   74 yIleGluLeuProPheAspSerCysAsnThrAlaArgThrArgSerLeuA 91
272 ATCCAGCTGTATTTTGTAAACAACAACGTGTTCATTCGTTTCATCA 321
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
   91 snProLysGlyValPheValSerThrThrValAlaIleSerPheHisPro 107
322 TTATTTGTTCCAAAGTTGATGCGCATGCGATGCGATGCGATGCGAT 371
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
   108 GlnPheValThrLysValAspArgAlaTyrArgIleGlnCysPheTyrMe 124
372 GGAAGCTGATTAACAGTAAATGTCAGATTCGATTCGATTCGATTCGAA 421
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
   124 tGluSerAspLysThrValSerThrGlnIleGluValSerAspLeuThr 141
422 CTGCTTTTCAACCTCAAAATGTCGCCGATGCCAGTATGCCGTTGAAATT 471
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
   141 hTrAlaPheGlnThrGlnValValPheMetProValCysLysTyrGlnIle 157
472 TTGGATGGTGACACCGGTCACAGATTCATTCGATTCATTCATTCGTC 521
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
   158 LeuAspGlyGlyProSerGlyGlnProIleGlnPheAlaThrIleGlyG 174
522 GCCAGTTTATCATTAATGACATGCGATTCGAAACCGTTGATCTTCT 571
   | : : : : : : : : : : : : : : : : : : : : : : : :
   174 nGlnValTyrHisLysTrpThrCysAspSerGluThrThrAspThrPheC 191
572 GCGCGGTTTCOCATTCCTGCTTTGTCATGATGTAACGATGATGCTGG 621
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
   191 ySaIaValAlaIleHisSerCysThrValAspArgGlyAsnGlyAspThrVal 207
622 GAAATTCATTAATGCTGATGCTGCTGATTAATTAATTAATTAATTAAT 671
   : : : : : : : : : : : : : : : : : : : : : : : : : :
   208 GlnIleLeuAsnGluGluGlyCysAlaLeuAspLysPheLeuLeuAsnAs 224
672 TTGGATATTCACACAGATTTAATGCGTGGCGCAAGACCTCACATTAACA 721
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
   224 nLeuGluTyrProThrAspLeuMetAlaGlyGlnGlnAlaAlaHisValTyrL 241
722 AATTCGCGATCGATCAGACCTTTTATCAATGCCAGATCAGATTAATACC 771
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
   241 yTyrAlaAspArgSerGlnLeuPheTyrGlnCysGlnIleSerIleThr 257
772 ATTTAAACAACCAATAGCAATGTGTCGACACCAATGTTCAAGAACACA 821
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
   258 IleLysAspProGlySerGluCysAlaArgProThrCysSerGluProG 274

```

```

822 AGCATTCGACGCTGTAAACAGCTGCT..... 849
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
274 nGlyPheGlyAlaValLysGlnAlaGlyAlaGlyAlaHisAlaAla 291
850 .....GCCGCA 855
291 IaIaProGlnAlaGlyValGluGluValGlnAlaAlaProValAlaIa 307
856 GCNAACCTGCTGACGT..... 873
308 AlaIaProValAlaIaIaProValAlaIaAlaIaIaIaIaIaIaIa 324
874 .....GCCACACTCTGTTACTC...AGAAAAGATCT. 903
324 lProArgAlaThrLeuAlaGlnLeuArgLeuLeuArgLysLysArgSer 341
904 ..GCAGACCGGAGATATGATTCATGATGTACGACCTGATATCAACCCCTT 951
341 heGlyGluAsnGluGlyIleLeuAspValAlaArgValGluIleAsnThrLeu 357
952 GAATTCAGCATGATATCAAGCTTTGCCAGTTGATTTAGTCACCGTGC 1001
358 AspIleMetGluGly.....AlaSerProSerAlaProGluAlaAla 372
1002 ACTTCTCACAATATGACACACCTGTAATACCTTGCTGCAGTACAAATG 1051
372 aLeuValSerGlu.....GluSerValArgArgAlaIaThrSerThrG 387
1052 GAATTCGATGTCACCACTTGGCTTCGAATGTTATGGGTTTAAGCATTT 1101
387 lLysCysLeuThrProIleGlyPheAlaSerPheLeuGlyIleGlyThr 403
1102 GCATTCGATTCGCGGTCATTTATACCATTCGTTTAATTTGGTCCA 1149
404 lLeValAlaThrAlaLeuSerAlaThrIlePheTyrValAlaArgPro 419

seq_name: SwissProt_38:YOH3_CAEEL

seq_documentation_block:
ID YOH3_CAEEL STANDARD; PRT; 495 AA.
AC 009276;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DI 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPONHEMICAL 55.1 KD PROTEIN C43C3.3 IN CHROMOSOME X.
GN C43C3.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilitida;
OC Rhabdilitina; Rhabdilitoidea; Rhabdilitidae; Peloderinae; Caenorhabdilitis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA SULSTON J.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
DR EMBL: 247067; CAAB7330.1; -
DR WORMPEP: C43C3.3; CE01525.
FT Hypothetical protein; Transmembrane; ATP-binding.
FT TRANSMEM 11 31 POTENTIAL.
FT NP_BIND 426 446 POTENTIAL.
FT NP_BIND 171 178 ATP (POTENTIAL).
SQ SEQUENCE 495 AA; 55068 MW; 09C77468 CRC32;

```


alignment_scores: Length: 340
 Quality: 108.00
 Ratio: 0.679
 Gaps: 19
 Percent Similarity: 46.765 Percent Identity: 22.647

alignment_block:

US-09-323-427-3 x YOH3_CAEEL ..

Align seg 1/1 to: YOH3_CAEEL from: 1 to: 495

```

91 ATGATGATGGACCACTTCATATACATCAATTTAAAT.....129
   :::::| | | | | | | | | | | | | | | | | | | | |
83 ValAspCysIleAlaAspSerPheThrValValLeuAsnLysSerAspPr 99
130 .....ACAGTAATGCATTCGAGGACATGTTATG 160
   :::::| | | | | | | | | | | | | | | | | | | | |
99 OGluValMetArgMetIleSerAsnProLysSerGlnProValValIleYrV 116
161 TGAAGGCTCTTATGATCAAGAGGTTGC.....CGTAATGATGAA 201
   || | | | | | | | | | | | | | | | | | | | |
116 aLYrGlyHisLysThrArgHisProCysGlyThrSerMetLysAspGlu 132
202 GGTGGAGCTCAAGTGGCCGCAATTCCTCCATTTGATTCGA...TCGAA 248
   || | | | | | | | | | | | | | | | | | | | |
133 LysGlyLeuThrAsnPheAsnLeuThrIleProTyrGlySerGlyCysAs 149
249 TGTGGCCGCTATCGAGATCTCGAATCCACGCGTATTTTGTACACAA 298
   :::::| | | | | | | | | | | | | | | | | | | | |
149 pValThrLeuThrAspLeu.....ProLysHisArgTyrAlaGluThrT 164
299 CTGTTGTCATTTCTGTTTCAT...CCATTAATTTTGTACCAAGATGATG 345
   || | | | | | | | | | | | | | | | | | | | |
164 hrValValLeuGluAspAsnAlaAspLeuSerPheGlyLysThrArg 180
346 GCATGAGTACATGCTTTTACATGGAAGCTGATAAAGACTTACTGTC 395
   :::::| | | | | | | | | | | | | | | | | | | | |
181 LeuAsnHisValPheCysLeuTyrThrArgAsnValLysThr..... 194
396 ACAGATGAGGATTCGAAATACACACTGCTTTCGA.....A 433
   || | | | | | | | | | | | | | | | | | | | |
195 .....IleArgPheSerAspValSerAsnGlyHisGluValIleAlaSerT 210
434 CTCGAATTTGCCGATGCCAGTATGCCGTTATGAAATTTTG..... 474
   || | | | | | | | | | | | | | | | | | | | |
210 hrGlyGlyLysProLysPro.....LysValGluMetLeuPheArgSer 224
475 ...GATGGTGACCAACCGGTCAACCA.....GTTCA 503
   || | | | | | | | | | | | | | | | | | | | |
225 ThrAspSerGlyLysThrLeuGlnAlaAlaArgGluAsnGluPheValG 241
504 ATTGCTATCATTTGTCAGCCA.....GTTTATCATTAATGAGCA.... 543
   || | | | | | | | | | | | | | | | | | | | |
241 upPhePheIleAlaLeuSerProAspSerAlaTyrHisGlyLysSerPro 258
544 .....TGCATTCGAAACCGTTGAT..... 564
258 ysgLysCysThrPheSerAspArgLysPheSerAlaProAspAlaLys 274
565 .....ACTTTCGCGCGGTTGTCATTCCTGCTTTGCGATGATGTTAA 608
   || | | | | | | | | | | | | | | | | | | | |
275 LysIleThrPhe.....ValGlnGlyGlyCysProValAsnGlyMetAs 289
609 CGGATACACGTGGAATTCCTAATGCTGATGATGTCCTTTGATTAAT 658
   |
289 nAsp..... 290
659 ATTGCTAATTAATTTGGAATATCCACAGATTTAATGCTGCCAAGAA 708
   :::::| | | | | | | | | | | | | | | | | | | | |
291 ..IleIleAspProLeuAlaAsnValAsnAspGlnIleTyrPheSerLys 306
709 GCTACAGTATCAAAATATGCGGATGATGATGATGATGATGATGATGAT 758
   :::::| | | | | | | | | | | | | | | | | | | | |
307 PheArgThrPheArgPheGlyAsnGlnSerThrValAlaPheValHisCysG 323

```

```

759 GATCATGATTACCATTAAGACCAATAGCAATGTTGACACCAAT 808
   :::::| | | | | | | | | | | | | | | | | | | | |
323 nValGlnValCysLeuLysLys.....AspGluCysSerLys...Thrc 337
809 GTTCAGAACCAACAGATTCGAGCTGTAAACAGGTGTGCGCGACGA 858
   || | | | | | | | | | | | | | | | | | | | |
337 ysrTyrLysLys.....ValSerAsp 343
859 AACCTGCTGCAGCTGCGCAACTTCGTTTACTCAAGAAAGATTCGACA 908
   :::::| | | | | | | | | | | | | | | | | | | | |
344 SerAsnLeuThrIleAlaGluArgLeuArgPheArgHisLysArgSer.... 358
909 ACCGAGAAATATCATGATGATGACGAACTGATATCAACACCCCTTGAAT 958
   || | | | | | | | | | | | | | | | | | | | |
359 .....IleThrAspLeuGluArgGlyThrThrArgSerAlaProT 372
959 GCGATGATTAATCAAGCTTTG 978
   :::::| | | | | | | | | | | | | | | | | | | | |
372 hrAspAspAsnGlySerLeu 378

```

seq_name: SwissProt_38:YAY3_SCHPO

seq_documentation_block:

ID YAY3_SCHPO STANDARD: PRT; 649 AA.

AC Q10211;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE HYPOTHETICAL 74.5 KD PROTEIN CAH3.03C IN CHROMOSOME 1.

GN SPACAH3.03C.

OS Schizosaccharomyces pombe (fission yeast).

OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomyces.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RA MURPHY L., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;

RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: Z69380; CAA93342.1; -

KW Hypothetical protein; Transmembrane.

FT TRANSMEM 265 285 POTENTIAL.

FT TRANSMEM 564 584 POTENTIAL.

FT TRANSMEM 626 646 POTENTIAL.

SO SEQUENCE 649 AA; 74488 MW; F04F8763 CRC32;

alignment_scores: Length: 403
 Quality: 97.50
 Ratio: 0.554
 Gaps: 23
 Percent Similarity: 43.672 Percent Identity: 22.333

alignment_block:

US-09-323-427-3 x YAY3_SCHPO ..

Align seg 1/1 to: YAY3_SCHPO from: 1 to: 649

```

73 CTCGAAGCTGAGCCAGAAATGATGTGGACCACTTCATTAACATCAA 122
   :::::| | | | | | | | | | | | | | | | | | | | |
36 LeuAspGlySerValGluMetMetCysTrpPro.....As 47
123 TTTTAATACAGTAATGCATTC.....GAGGAC 151

```

```

47  npheserProserIlePheAlaArgIleLeuAspAlaArgAlaGlyH 64
152 ATGTTATGTAAGAGTCTTATGATCAAGAAGTGGCCATGATGATA 201
64  IspheSerIleThrProIle...GluGlnThrSerCysLys... 76
202 GGTGACAGTCAAGTTGGCCGAATTTGACCTTCATTCATTCATGCAATGT 251
77  .....GlnMetIleGluProSerThrAsnIle 85
252 TGCGGTACAGCATCTGTAATCCAGCTGATATTTGTAACAACAAGT 301
85  eLeuHisThrLysPheThrSerGluArgGlyValLeu...ArgL 99
302 TTGTCATTCGTTTCAT.....CCATTATTTGTT 330
99  euleuAspPhePheHisArgProThrGluAspTyrGluProLeuIleuThrPro 115
331 ACCAAGTTGATCGTGA.....TATCGAGT 356
116 TrpLeuIleArgValSerCysIleArgGlyThrSerArgIleLysLe 132
357 ACAATGCTTTACATGAGAGTGAATAACAGTTAGTCCACAGATTGAG 406
132 uGluCysPheProAlaLeuAspTyrAlaArgGlnSerHisGluThrArgV 149
407 TATGGAATGACACA...ACTGCTTTTCAACATGATGTCGCCAGTCCA 453
149 alSerLysIleThrGluAsnTyrTyrGlnAlaGluPheValProAlaSer 165
454 GTATGCCCTTATGAATTTGATGGTGACCAACCGGTCAA..... 495
166 GlyAspProLysTyrIleLeuAspCysValProSerGlyAspGlnLeuL 182
496 .....CCAGTTCAATTTGCTATCATTTGCTAC 523
182 sIleAspLeuGluLeuIleTyrProAlaGlnHisLeuIleGluLysG 199
524 CAGTTTATCATAAATGACATGCGATTCGTAACCGTTGATGATTTCGC 573
199 LyValIleSerTyrLeuGluLeuGluGlnGluGlnIleThrPhe... 214
574 GCGGTGTGCATTCCTGTTGTGATGATGTAACGCTGATGACTGTGA 623
215 .....IlePheArgGlnGluLysLeuGlyProAsnValas 226
624 AATTTCAATGCTGATGATGCTGCTTGTGATAATATTGCTAATAATAT 673
226 pTyrValThrProAsn.....LeuValAspLysL 236
674 TGAATATCCACAGATTTAATGCTGGCCAGAAGCTCAGTATACAAA 723
236 euGluAspSerThrLysArgTyrTrpArgAlaThrPleGlnGlnCysVal 252
724 TATCGGATGATCAGACGCTTTCTATCAA.....TGCCAGATCAG 764
253 TyrThrGlyArgTyrArgGluPheValGlnArgAsnAlaLeuThrLeu 269
765 TATTACCTTAAAGACCAATAGCGAATGTGT...CGACCACAATGT 811
269 sLeuLeuIleTyrGluProThrGlyAlaValIleAlaSerProThrPhe 286
812 CAGAACCAAGATTCGAGCTGTTAAACA..... 843
286 erLeuProGluAspLeuGlyValAlaArgAsnTrpAspTyrArgPheThr 302
844 .....GGTGTGCCGACGAAAAACGCTGCGACCTGCGCAACTT... 882
303 TrpLeuArgAspSerAlaPheThrIleTyrAlaLeuAlaGlnLeuGlyPh 319
883 .....CGTTACTACAAG 895

```

```

319  eArgAlaGluAlaValGluTyrMetSerPheIleTyrHisValLeuLysL 336
896  AAAGATCGCAGACCGGAG...AATATCATGATGTGACAGCATGATATC 942
336  yslYsAsnLysAspGlyGlyIleAsnIleVal..... 346
943  AACACCCCTTGAATATACGATGATATCAAGCTTTG...CCAGTTGATTT 989
347  ....TyrSerIleHisGlyAspSerGlnAsnLeuGluValGluLe 361
990  ACGTCAC...CGTGCACCTTCTGCAACATATGAGACAACTGTATCTGTG 1036
361  uThrHisLeuArgGly.....TyrTyrAsnSerHisProValArgIle. 375
1037  CTCGAGTACAATAATGAGATGCGATGTCACCATTTGGCTGTGCAATGTT 1086
376  .....GlyAsnAlaValAlaHisHisLeuGlnLeuAspIleT 388
1087  ATGGG 1091
388  yGly 389

seq_name: SwissProt_38:APMU_PIG
seq_documentation_block:
ID  APMU_PIG          STANDARD:      PRT:  1150  AA.
AC  P12021;
DT  01-OCT-1989 (Rel. 12, Created)
DT  01-DEC-1992 (Rel. 24, Last sequence update)
DT  15-JUL-1998 (Rel. 36, Last annotation update)
DE  APOMUCCIN (MUCIN CORE PROTEIN) (FRAGMENT).
OS  Sus scrofa (Pig).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
   OC  Eutheria; Cetartiodactyla; Suidae; Sus.
RP  [1]
RC  SEQUENCE FROM N.A.
RC  TISSUE=SUBMAXILLARY GLAND;
RA  ECKHARDT A.E., TIMPTE C.S., ABERNETHY J.L., ZHAO Y., HILL R.L.;
RT  "Porcine submaxillary mucin contains a cysteine-rich,
RT  carboxyl-terminal domain in addition to a highly repetitive,
RT  glycosylated domain.";
RN  J. Biol. Chem. 265:9678-9686(1991).
RN  [2]
RP  SEQUENCE OF 1-503 FROM N.A.
RC  TISSUE=SUBMAXILLARY GLAND;
RX  MEDLINE: 88087170.
RA  TIMPTE C.S., ECKHARDT A.E., ABERNETHY J.L., HILL R.L.;
RT  "Porcine submaxillary gland apomucin contains tandemly repeated,
RT  identical sequences of 81 residues.";
RN  J. Biol. Chem. 263:1081-1088(1988).
RN  [3]
RP  SEQUENCE OF 45-80.
RC  TISSUE=SUBMAXILLARY GLAND;
RX  MEDLINE: 87280230.
RA  ECKHARDT A.E., TIMPTE C.S., ABERNETHY J.L., TOWMADGE A.,
RA  JOHNSON W.C. JR., HILL R.L.;
RT  "Structural properties of porcine submaxillary gland apomucin.";
RN  J. Biol. Chem. 263:11339-11344(1987).
RN  [4]
RP  CARBOHYDRATE-BINDING SITES, AND SEQUENCE OF 45-125.
RC  TISSUE=SUBMAXILLARY GLAND;
RX  MEDLINE: 97248516.
RA  GERKEN T.A., OWENS C.L., PASUMARTHY M.;
RT  "Determination of the site-specific O-glycosylation pattern of the
RT  porcine submaxillary mucin tandem repeat glycopeptide. Model proposed
RT  for the polypeptide:galnac transferase peptide binding site.";
RN  J. Biol. Chem. 272:9709-9719(1997).
CC  -I- FUNCTION: APOMUCCIN IS PART OF MUCIN, THE MAJOR GLYCOPROTEIN
CC  SYNTHESIZED AND SECRETED BY MUCOUS CELLS OF THE SUBMAXILLARY
CC  GLAND. ITS HIGHLY VISCOUS AQUEOUS SOLUTIONS SERVE TO LUBRICATE
CC  THE ORAL CAVITY AND TO PROTECT IT FROM THE EXTERNAL
CC  ENVIRONMENT.

```

```

FT CARBOHYD 114 114
FT CARBOHYD 117 117
FT CARBOHYD 123 123
FT CARBOHYD 124 124
FT CARBOHYD 124 124
FT CARBOHYD 418 418
FT CARBOHYD 547 547
FT CARBOHYD 917 917
FT CARBOHYD 985 985
FT CARBOHYD 1002 1002
FT CARBOHYD 1068 1068
SQ SEQUENCE 1150 AA; 109615 MW; F7C55CCE CRC32;

alignment_scores:
    Quality: 98.00      Length: 221
    Ratio: 1.089        Gaps: 12
    Percent Similarity: 40.724      Percent Identity: 23.982

alignment_block:
US-09-323-427-3 x APMU_PIG ..

Align seg 1/1 to: APMU_PIG from: 1 to: 1150

382 AAAACAGTTATGTCACACAGATTAGATCTGA...ATCACAGCTGTT 428
   ||| :::::|||||: ::|||::: |||||
892 LysGluIlethrIaSerProLysValSerSerProGluThrIlaGl 908
   429 TCAACTCA.....A 439
   |||:::
908 yalathrGluaspGlnGluasnGlysthGlyCysProIaProL 925
   440 TTTCCTCCGATGCCAGTATGCCGTTATGAAATTTGGATGGTGACCAC 489
   :: ||| |||||::: |||||
925 eurProProProValCysHis.....GlyProIeu 935
   490 GGTCAACCACTCAATTGCTATCATTTGGTCAGGCCAGTTATCAATAATG 539
   |||::: ::::: |||||
936 GtGluGluIuLysSerProGlyAspValThrIlaIncysHISLysC 952
   540 GACATG...GATTGCAACCCGTGAT.....ACTTCTGGGGG 577
   |||||: :::::||||| |||
952 sthrcyssthGluIaLysThrValAspCysLysProLysGluCysPro 969
   578 TTCTCCATTCCTGCTTTGCGATGAT.....GGTAA 609
   969 eurProthrcysLysThrGlyGluIaArgLeuIleLysPheLysala 985
   610 GGTGATACTGTGCAATCTTAATGCTGATGAGATGCTCTTGATAATA 659
   986 AspThrCysCysGluIle.....GlyHisCysGluLysArgThrC 999
   660 TTTCGTAATAATTTGGATATATCCACAGATTATGGCTGGCAAGAAG 709
   999 sIeuPheasnAsnThrAspTyr.....1006
   710 CTCACGTATCAAAATATATGCGATGATCACAGCTTTCTATCAATGCC 759
   1007 .....Glu 1007

760 ATCAGATATTCACATTAAAGAACAATAGCAATGTGTGCACCAATG 809
   1008 ValGlySerSerPheAspAspProAsnAsnProCysValThrTyrSec 1024
   810 TTGAGAACCAAGAGATTGGAGCTGTAA.....AACA 843
   1024 s...GlnAsnThrGlyPheThrIaIaValGlnAsnCysProLysGln 1040
   844 GGTGTGTCCG.....ACCAAACTGCTGC... 869
   1040 hrrTrpCysIaGluGluAspArgValTyrAspSerLysGlnCysCys 1056
   870 AGCTGGCCACTTCGTTTACTCAAGAAAGATCTGCAGAACCGGAGATA 919
   ::|||

```

1057 ThrCysLysSerSerCysLysProSerProValAsnValThrValArgTyr 1073
 920 TCATGATGATGACG 932
 1073 rAsnGlyCysThr 1077

seq_name: SwissProt_38:GALU_PSEAE

seq_documentation_block:
 ID GALU_PSEAE STANDARD: PRT: 279 AA.
 AC 059633:
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE UTP--GLUCOSE-1-PHOSPHATE URIDYLTRANSFERASE (EC 2.7.7.9) (UDP-GLUCOSE
 DE PYROPHOSPHORYLASE) (UDPGP) (ALPHA-D-GLUCOSYL-1-PHOSPHATE
 DE URIDYLTRANSFERASE) (URIDINE DIPHOSPHOGLUCOSE PYROPHOSPHORYLASE).
 GN GALU.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
 OC Pseudomonas.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 10145;
 RA CHANG H., LEE C., PENG H.;
 RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY PLAY A ROLE IN STATIONARY PHASE SURVIVAL (BY
 CC SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: UTP + ALPHA-D-GLUCOSE 1-PHOSPHATE =
 CC PYROPHOSPHATE + UDP-GLUCOSE.
 CC -1- SIMILARITY: BELONGS TO THE PROKARYOTIC UDGP FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U03751; AAB01486.1; -
 CC DR PPM: PF00483; NTP_transferase; 1.
 CC DR Transferase; Kinase; Nucleotidyltransferase.
 CC KW SEQUENCE 279 AA; 30939 MW; D967F501 CRC32;
 CC SQ

alignment_scores:
 Quality: 96.00 Length: 236
 Ratio: 0.807 Gaps: 12
 Percent Similarity: 50.424 Percent Identity: 20.763

us-align_block:
 us-09-323-427-3 x GALU_PSEAE ..

Align seg 1/1 to: GALU_PSEAE from: 1 to: 279

55 ATTCGGTTCACAT.....GGTGTGAGGTGAGCC 86
 ::::::::::||| ||| ||| ||| |||
 28 LeuProValValAsnLysProLeuIleGlnTyrGlyVal.....GluG1 42
 87 AGAATTCGATGTGACCACTTCATATACATC...AATTTTATATAC 133
 ::::::::::||| ::::::::::||| :::::
 42 uAlaLeuAspAlaGlyLeuAsnGlnIleSerIleValThrGlyArgGlyL 59
 134 GTATGATTCGAGAGACATGTTATGTAAGGCTTTATGATCAGCA 183
 ::::::::::||| ||| :::::
 59 ySAIgaAlaLeuGluAspHisPheAspIleSer.....TyrGluLeuGlu 73
 184 GGT...TGCCGTAATGATGAAGTGTGACGACGACGTAAGTCCGGAATTC 230
 ::::::::::||| :::::
 74 AsnGlnIleLysGlyThrAspLysGluLysTyrLeuValGlyIleArgGly 90
 231 TCATTTGATTCATGCAATGTTGGCGGTACAGCATCTCTGATCCACGTG 280

90 sLeuLeuAspGlyCysSerPheSerTyrThrArgGlnThrGlnIleMetLysG 107
 281 GTATT.....TTGTACACACACTGTTGCATTCGTTGCATCCCA 321
 ||::::::::::::::::::|||
 107 lYleGlyHisAlaIleLeuThrGlyArgProLeuIleGlyAspLeuPro 123
 322 TTATTTGTACCAAGTTGATCGATGCAATGAGTACAACTGTTTACAT 371
 ::::::::::||| |||
 124 PheAlaValValLeuAlaAspAspLeu.....CysValAsnLe 136
 372 GGAAGCTGATTAACAGCTTACTGCACAGATTGAGTATTCGAA..... 414
 ::::::::::||| ::::::::::|||
 136 uGlyGlyAspGlyValLeuThrGlnMetValLysLeuTyrGlnLysTyr 153
 415 ..ATCACACGCGTTTTCAAACGTAATTCGCCGACGCGTATGCGGT 462
 |||
 153 rGcysThrIleValAlaValMetGluValAsnProThrGluThrAsnLys 169
 463 TATGAAATTTTGGATGCT..... 480
 |||
 170 TyrGlyValIleAlaGlyAspAspIleGlyAspGlyLeuIleArgValAr 186
 481GGACCAACCGGTCAACACAGTTCAATTGCTA 511
 ::::::::::|||
 186 gAspMetValGluLysProAlaProGluAspAlaProSerAsnLeuAlaI 203
 512 TCATTTGTCACCCAGTTTATCATTAATGACATGCGATTCGAAACGCTT 561
 |||
 203 lIleGlyArgTyrIleLeu.....ThrPro 211
 562 GATACCTTTCGCGCGTGTTCATTCCTGCTTTCGATGATGATGATGACG 611
 |||
 212 AspIlePheLysLeuIleGluGluThr.....GluProGlyLysG1 225
 612 TGATACCTGTGGAATTCCTAAATGCT.....GATGAT 643
 ::::::::::|||
 225 yGlyGluIleGlnIleThrAspAlaLeuLysGlnAlaLysAspLysC 242
 644 GTGCTCTT 651
 |||
 242 yValIle 244

seq_name: SwissProt_38:AMD2_XENLA

seq_documentation_block:
 ID AMD2_XENLA STANDARD: PRT: 875 AA.
 AC P12850;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PEPTIDYL-GLYCINE ALPHA-AMIDATING MONOOXYGENASE II PRECURSOR
 DE (EC 1.14.17.3) (PEPTIDE C-TERMINAL ALPHA-AMIDATING ENZYME II) (AE-II).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
 OC Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae;
 OC Xenopus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SKIN.
 RX MEDLINE: 88134244.
 RA OHSHYE K., KIRANO K., WADA Y., FUCHIMURA K., TANAKA S., MIZUNO K.,
 RA MATSUO H.;
 RT "Cloning of cDNA encoding a new peptide C-terminal alpha-amidating
 RT enzyme having a putative membrane-spanning domain from Xenopus laevis
 RT skin."
 RL Biochem. Biophys. Res. Commun. 150:1275-1281(1988).
 CC -1- FUNCTION: C-TERMINAL ALPHA-AMIDATION OF BIOLOGICAL PEPTIDES.
 CC -1- CATALYTIC ACTIVITY: PEPTIDYLGLYCINE + ASCORBATE + O(2) =
 CC PEPTIDYL(2-HYDROXYGLYCINE) + DEHYDROASCORBATE + H(2)O
 CC (THE PRODUCT IS UNSTABLE AND DISMUTATES TO GLYOXYLATE AND THE
 CC CORRESPONDING DESGLYCINE PEPTIDE AMIDE).
 CC -1- COFACTOR: COPPER, AND ASCORBATE.

```

CC -1- SIMILARITY: BELONGS TO THE COPPER TYPE II, ASCORBATE-DEPENDENT
CC MONOOXYGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M19032; AAA49667.1; -
CC PIR: A27715; ORXLA2.
CC PROSITE: PS00084; CU2_MONOOXYGENASE_1; 1.
CC DR PROSITE: PS00085; CU2_MONOOXYGENASE_2; 1.
CC DR PFAM: PF01082; Cu2_monooxygen; 1.
CC DR PFAM: PF01436; NHL; 4.
CC KW Oxidoreductase; Monooxygenase; Copper; vitamin C; Transmembrane;
CC Glycoprotein; Signal.
CC SIGNL 1 39
CC FT CHAIN 40 875
CC FT DOMAIN 40 763
CC FT TRANSMEM 764 787
CC FT DOMAIN 788 875
CC FT DOMAIN 358 364
CC FT DOMAIN 387 390
CC FT DOMAIN 852 856
CC FT CARBOHYD 465 465
CC FT CARBOHYD 662 662
CC FT CARBOHYD 743 743
CC SQ SEQUENCE 875 AA; 97084 MW; 06EEF456 CRC32;

```

```

alignment_scores:
  Quality: 94.50      Length: 232
  Ratio: 0.945      Gaps: 12
  Percent Similarity: 43.103      Percent Identity: 21.552

```

```

alignment_block:
US-09-323-427-3 x AMD2_XENLA ..

```

```

Align seg 1/1 to: AMD2_XENLA from: 1 to: 875

```

```

97 TGTGACCACTTCATATACATCATTTTATACACAGTATGATTCGA 146
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
530 CysGlnProThrAspValAlaValAspProIleHrGlyAsnPhlePheVa 546
147 AGACATGTTTATGTGAAGGCTTTATGATCAGAAGGTGCGGTATG 196
   :: |||
546 IAlAspGlyTyr..... 550
197 ATGAAGTGGACGTCACAGTTCGCCGAAATTCACCTTCATTGATTCATGC 246
551 ..... 551
247 AATGTGGCGGTACACGATCTCTGATTCACGCGTATTTTGTGA..... 291
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
552 Asn...SerAlGileMetGlnPheSerProAsnGlyMetPheIleMetGcl 567
291 ..... 291
567 nTrpGlyGluGlnThrSerSerAsnLeuProArgProGlyGlnPheArgi 584
292 ..... ACACAACTGTTGTCATTTGCTTTCATTCATTCATTTGTTACC 333
584 leProHISerLeuThrMetIleSerAspGlnGlyGlnLeuGlyValAla 600
334 AAAGTTCGTCGATCGAGTCAATGCTTTTATGATGACGAGCGATGATA 383
601 AspArgGlyAsnGly...ArgIleGlnCysPheHisAlaValSthGlyGI 616
384 AACAGTTAGTGCACAGATTGAGTATCTGAATC...ACAACGTGCTTTTC 430

```

```

seq_name: SwissProt_38:YLK3_CAEEL
seq_documentation_block:
ID YLK3_CAEEL STANDARD; PRT; 1895 AA.
AC P41951;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE PUTATIVE SERINE/THREONINE-PROTEIN KINASE D1044.3 IN CHROMOSOME III
DE (EC 2.7.1.-).
GN D1044.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA PAULEY A., WATERSTON R.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: STRONG, TO ZC84.1.
CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
CC PROTEIN KINASES.
CC -1- SIMILARITY: CONTAINS ? EGF-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U00065; AAA50735.1; -
CC DR HSSP: O63450; 1A06.
CC DR MORMEP: D1044.3; CE01206.
CC DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
CC DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
CC DR PFAM: PF00069; PKINASE; 2.
CC KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;
CC ATP-binding; EGF-like domain.
CC FT DOMAIN 431 703
CC FT NP_BIND 437 445
CC FT BINDING 461 461
CC FT ACT_SITE 569 569

```


DR EMBL: M36192; AAA30657.1; -
 DR HSP; P00214; 1FTC.
 DR PROSITE: PS01208; VWF; 1.
 DR PROSITE: PS01185; CTCK_1; 1.
 DR PROSITE: PS01225; CTCK_2; 1.
 DR PFM; PF00007; Cys_knot; 1.
 KW Repeat; Glycoprotein.
 FT DOMAIN 61 158 3 X 11 AA REPEATS.
 FT REPEAT 61 158 1.
 FT REPEAT 112 122 2.
 FT REPEAT 148 158 3.
 FT DOMAIN 338 404 VWF.
 FT DOMAIN 471 535 CTCK.
 FT SIMILAR 1 282 TO PORCINE APOMUCIN.
 FT DISULFID 471 518 BY SIMILARITY.
 FT DISULFID 485 532 BY SIMILARITY.
 FT DISULFID 494 548 BY SIMILARITY.
 FT DISULFID 498 550 BY SIMILARITY.
 FT DISULFID 2 554 BY SIMILARITY.
 FT CARBOHYD 28 28 POTENTIAL.
 FT CARBOHYD 394 394 POTENTIAL.
 FT CARBOHYD 477 477 POTENTIAL.
 SO SEQUENCE 563 AA; 58913 MW; BB7D4189 CRC32;

alignment_scores:
 Quality: 90.00 Length: 345
 Ratio: 0.612 Gaps: 18
 Percent Similarity: 42.609 Percent Identity: 20.290

alignment_block:
 US-09-323-427-3 x MUCS_BOVIN ..

Align seg 1/1 to: MUCS_BOVIN from: 1 to: 563

```

61 GTTGACATCTGTGCGAAGTGGAGCCGAATGATGACCAACTTC 110
   :::::::::::::::::::: ||| |||
192 lIeGIhnrGlyIeIhnrGlyhnrGlySer.....GlyhnrIhSe 205
   :::::::::::::::::::: ||| |||
111 AATAACATCAATTTTATACAGCT...AATGATCGAAGACATGTTT 157
   :::::::::::::::::::: ||| |||
205 rSerProGlyGlyPheAsnIaGluIaIhnrPhelyGluIhVala 222
   :::::::::::::::::::: ||| |||
158 ATGTGAAAGCTTTTATGATCAGAGAGTGGCCATATGATGAAGTGA 207
   :::::::::::::::::::: ||| |||
222 rGhnrIhnrGlyhnrGlyIeUeSerGlyhnrIhnrGlyhnrGly 238
   :::::::::::::::::::: ||| |||
208 CGTCAAGTTCGCGAATTTTCATTTGATTCATGATCGATGTTGGCG 257
   :::::::::::::::::::: ||| |||
239 ThnrThVal.....IlePro...GluSerSerAsnIhnrGlyh 250
   :::::::::::::::::::: ||| |||
258 TACAGATCTCTGTGAATCCAGCTGGTATTGTAACAACAAGTGTGCA 307
   :::::::::::::::::::: ||| |||
250 rSerThGlyValGlyhnrGln.....ThSerThAlaVal 263
   :::::::::::::::::::: ||| |||
308 TTTCGTTTCATCCATTATTGTTAC..... 333
   :::::::::::::::::::: ||| |||
263 alSerGlyhnrGlyhnrGlyValSerGluSerSerProGlyThSer 279
   :::::::::::::::::::: ||| |||
333 ..... 333
280 LysGluAlaSerGluThnrIhnrGlyProGlyIleSerThnrGlySe 296
   :::::::::::::::::::: ||| |||
334 ..... 367
296 rIhnrSerIySerAsnArgIleThnrIhnrSerSerArgIlePro....T 311
   :::::::::::::::::::: ||| |||
368 ACATGGAAGCTGATAAACAGTTAGTCACAGATGAGGTATCTGAATC 417
   :::::::::::::::::::: ||| |||
311 yrProGluThnrIhnrValaIaIhnrGlyGluGln.....ThGluThr 326
   :::::::::::::::::::: ||| |||
418 ACAACGTCTTTCAACTCAATTTGCCGATGCCAGTATGCCGTTATGA 467
   :::::::::::::::::::: ||| |||

```

```

327 LysThrGlyCysThnrIhnrSerIeuProProProAlaCysTyr..... 341
468 AATTTGGATGTGAGCAACCGGTCAACAGTTCATTTGCTATGATG 517
   ||||| ||||| :::: |||
342 .....GlyProIeuGlyGlyIySerProGlyAspIleT 354
   ||||| ||||| ||||| |||||
518 GTCAGCCAGTTTATCATTAATGACATGC...GATTCGAACCGTTGAT 564
   ||||| ||||| ||||| |||||
354 rPhrIaAsnCysIhnrGlySerThCysThrAspIaGluThrValaSp 370
   ||||| ||||| ||||| |||||
565 .....ACTTTCGGCGGTTGCCATTCCTGCTTTTCATGATG 605
   ||| ||| :::: |||
371 CysIySerIeuGlySerProSerProProIhnrCysIySerProGlu 386
   ||| ||| ||||| |||||
606 TAACGGTACTGTGGAAATCTAAATCGTATGATGCTGCTTGAATA 655
   ||||| ||||| ||||| |||||
387 .....ArgIeuValIySerPheIyAsnAspIhnrCysGluIle 401
   ||||| ||||| ||||| |||||
656 AATAT.....TTGCTAAATGATTTGGAAATGCCAACA 687
   ||| ||| ||||| |||||
401 IatyrCysGluProArgThrCysIeuPheAsnAsnAspTyr..... 415
   ||| ||| ||||| |||||
688 GATTTATGCGTGGCCAGAACGTCACGTATACAAATATGCGGATGATC 717
   ||| ||| ||||| |||||
415 ..... 415
738 ACAGCTTTCTATCAATGCCAGATCATATTACATTAAGAACAAATA 787
   :::::::::::::::::::: ||||| |||||
416 .....GluValGlyAlaSerPheAlaAspProIyAs 426
   ||||| ||||| ||||| |||||
788 GCGAATGTGTTGACCAACAATGTTTCAGAACCCAGAGATTTGGAGTGT 837
   :::::::::::::::::::: ||||| |||||
426 snProCysIleSerTyrSerCysIhnrAsn...ThrGlyPheValaIaVal 441
   ||||| ||||| ||||| |||||
838 AA.....AACAGTGTGCGCC..... 854
   ||||| ||||| ||||| |||||
442 ValGlnAspCysProIyGlnIhnrIhnrCysValGluGluAspArgValIy 458
   ||||| ||||| ||||| |||||
855 .....AGCAAAACCTGCTGCAGCTGCCCAACTGTTTACTCAAGAAAGA 900
   ||||| ||||| ||||| |||||
458 rAspSerThrIySerCys.....TyrThrCysIySerProT 470
   ||||| ||||| ||||| |||||
901 TCTGCAGAACCGAGATATCATTTGATGTACGAC 935
   ||||| ||||| ||||| |||||
470 yrcysArgSerSerSerValAsnValThrValaSn 481
   ||||| ||||| ||||| |||||
seq_name: swissprot_38:NRK1_YEAST
seq_documentation_block:
ID NRK1_YEAST STANDARD; PRT; 1080 AA.
AC P38692;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE SERINE/THREONINE-PROTEIN KINASE NRK1 (EC 2.7.1.-) (N-RICH KINASE 1).
GN NRK1 OR YHR102W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DC-5;
RA FUKAMI Y.;
RL Submitted (MAY-1994) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RX MEDLINE: 94378003.
RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DOVER J.,
RA DU Z., FAVELLO A., FULTON L., GATTUNG S., GEISL C., KIRSTEN J.,
RA KUCABA T., HILLIER L., JIER M., JOHNSTON L., LANGSTON Y.,
RA LATREILLE P., LOUIS E.J., MACRI C., MARDIS E., MENEZES S., MOUSER L.,
RA NHAN M., RIFKIN L., RILES L., ST PETER H., TREVASKIS E., VAUGHAN K.,

```


RA VIGNATI D., WILCOX L., WOHLMAN P., WATERSTON R., WILSON R.,
 RA VAUDIN M.;
 RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
 RT VIII.";
 RL Science 265:2077-2082(1994).
 CC -I- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
 CC PROTEIN KINASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D29980; BA006250.1; -;
 DR EMBL; U000059; AB06860.1; -;
 DR PIR; S48944; S48944.
 DR HSSP; P24941; IAO1.
 DR SGD; L0001276; NRK1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP.1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST.1.
 DR PFAM; PF00069; PKINASE.1.
 KW Transferase; Serine/threonine-protein kinase; ATP-binding.
 FT DOMAIN 23 276 PROTEIN KINASE
 FT NP_BIND 29 37 ATP (BY SIMILARITY).
 FT BINDING 52 52 ATP (BY SIMILARITY).
 FT ACT_SITE 144 144 BY SIMILARITY.
 SO SEQUENCE 1080 AA; 117061 MW; E1668B6F CRC32;

alignment_scores:
 Quality: 90.00 Length: 209
 Ratio: 0.833 Gaps: 6
 Percent Similarity: 51.675 Percent Identity: 21.531

alignment_block:
 US-09-323-427-3 x NRK1_YEAST ..

Align seg 1/1 to: NRK1_YEAST from: 1 to: 1080

```

430 CAACACTCAATGTCCTCCGATGCGATGATCCGTTATGAATTTGGATG 479
      :::::::::::::::::::: ||| ||| ||| |||
200 LysValAspIleThrSerLeuGlyIleThrThrTyrGluIleAlaThrG 216
      :::::::::::::::::::: ||| ||| ||| |||
480 TGGACCA.....ACCGGTCAACGATTCATTTGCTA 511
      :::::::::::::::::::: ||| ||| ||| |||
216 ysnpPropotyrcysaspyValGluAlaLeuArgAlaMetGluLeuIle 233
      :::::::::::::::::::: ||| ||| ||| |||
512 TCATTCGTCAGCCGATTATCATTAATGACATGATTCGTAACCGTT 561
      :::::::::::::::::::: ||| ||| ||| |||
233 IeyserryssProProArgLeuGluAspArgSerTyrSerThrSerLeu 249
      :::::::::::::::::::: ||| ||| ||| |||
562 GATCTTTTCGCGCGGTGTCATTCCTGCTTTCGATGATGTAACGG 611
      :::::::::::::::::::: ||| ||| ||| |||
250 LysGluPheIleAlaLeu.....CysLeuAspGluAspProlys.. 262
      :::::::::::::::::::: ||| ||| ||| |||
612 TGATCTGTGGAATTCCTAAATCGTCGATGCTCTGATTAATATT 661
      :::::::::::::::::::: ||| ||| ||| |||
263 .....GluArgLeuSerAlaAspAspLeuLeuLysSerLysPheI 276
      :::::::::::::::::::: ||| ||| ||| |||
662 TGCTAATAATTTGGAATATCCACAGATTTAATGCTGGCCCAAGAAGCT 711
      :::::::::::::::::::: ||| ||| ||| |||
276 IeArgAlaIleLysAlaIleThrProThrSerIleLeuLysGluLeuIle 292
      :::::::::::::::::::: ||| ||| ||| |||
712 CACGATATCAATATGCGATGATGATGATGATGATGATGATGATGAT 761
      :::::::::::::::::::: ||| ||| ||| |||
293 ArgTyrLeuLeuPheArgAspLysAsnLysTyrLysIleGluI 309
      :::::::::::::::::::: ||| ||| ||| |||
762 CAGTATTACCATTTAAGAACAATAGCGAATGTGTCGACACCAATGTT 811
      :::::::::::::::::::: ||| ||| ||| |||

```

```

309 yserIleProGluAsnGluProSerLysProSerGluAlaProLysPro 326
      :::::::::::::::::::: ||| ||| ||| |||
812 CAGACCAACAGAGATTGCGAGCTGTAAACAGGTGTCGCCAGAGAAA 861
      :::::::::::::::::::: ||| ||| ||| |||
326 erGln.....AsnGlyGlyLysAspGluAlaGln 335
      :::::::::::::::::::: ||| ||| ||| |||
862 CCTGTCGACGTCGCAACTTCCTTACTCAAGAAAAGATCTGCAGAAC 911
      :::::::::::::::::::: ||| ||| ||| |||
336 LysSerIleAlaSerAsnAspAsnGluIleLysArgValAsn..... 349
      :::::::::::::::::::: ||| ||| ||| |||
912 GCGAATATCATGATGATGATGATGATGATGATGATGATGATGATG 961
      :::::::::::::::::::: ||| ||| ||| |||
350 GlnGlyAspValGluMetLysTyrAspPheAspSerLeuSerSerA 366
      :::::::::::::::::::: ||| ||| ||| |||
962 AT.....GATAATCAAGCTTGGCAGTGTATTA 990
      :::::::::::::::::::: ||| ||| ||| |||
366 spyrIleIleGluAsnAsnIleAsnLeuAspAlaLeuAlaGluAspAsn 382
      :::::::::::::::::::: ||| ||| ||| |||
991 CGTCACCGTCGCTCTGCAACATTAAT 1017
      :::::::::::::::::::: ||| ||| ||| |||
383 AsnGluThrPalaIleThrAlaGlnHisAsp 391
      :::::::::::::::::::: ||| ||| ||| |||

```

seq.name: SwissProt_38:TUD_DROME

seq_documentation_block:
 ID TUD_DROME STANDARD; PRT; 2515 AA.

```

AC P25823;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE MATERNAL TUDOR PROTEIN.
TUD.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP MEDLINE: 92038995.
RX SEQUENCE FROM N.A.
RA "Tudor, a posterior-group gene of Drosophila melanogaster, encodes a
RT novel protein and an mRNA localized during mid-oogenesis.";
RL Genes Dev. 5:2060-2070(1991).
CC -I- FUNCTION: REQUIRED DURING OOGENESIS FOR THE FORMATION OF
CC PRIMORDIAL GERM CELLS AND FOR NORMAL ABDOMINAL SEGMENTATION.
CC -I- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT THE LIFE CYCLE.
CC -I- REGION OF THE DEVELOPING OOCYTE DURING THE EARLY TO MIDDLE STAGES
CC OF OOGENESIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62420; CAA44286.1; -;
DR PIR; A41519; A41519.
DR FLYBASE; FBgn0003891; tud.
DR PFAM; PF00567; TUDOR; 10.
KW Developmental protein.
SQ SEQUENCE 2515 AA; 285236 MW; 1C381DB3 CRC32;

```

alignment_scores:
 Quality: 90.00 Length: 367
 Ratio: 0.552 Gaps: 16
 Percent Similarity: 44.414 Percent Identity: 19.074

alignment_block:
 US-09-323-427-3 x TUD_DROME ..

Align seg 1/1 to: TUD_PROME from: 1 to: 2515

```

46 TCCTATTGATCCGGTTGACAAATGGTGCAGAGGTGACGACGAATTTGA 95
   ||| :||||| ||| :||||| :||||| |||
1904 SerLeuInleuPro...AspAlaTyrIleSerThrSerProGluVal 1919
   ||| :||||| :||||| :||||| :|||||
96 ATGTGACCACTTTCATATACA.....ATCAATTTTAATACAC 133
   ||| :||||| :||||| :||||| :|||||
1919 uAlaLysPheAlaGluLeuThrGlyGluGlyLeuValPheThrThrG 1936
   ||| :||||| :||||| :||||| :|||||
134 GTATGCAATTCGAAGA.....CATGTTATGTGAAGGCTTTTATGAT 177
   ||| :||||| :||||| :||||| :|||||
1936 InleuLeuLysProGlyGlnAspHisValThrIleAspLeuLeuAsp 1952
   ||| :||||| :||||| :||||| :|||||
178 CAGAAGGTTGCCGTATGATGAAGGAGCAGTCAAGTCCGGAATTTTC 227
   ||| :||||| :||||| :||||| :|||||
1953 .....GlyGlnsniIleIleAspArgLe 1960
   ||| :||||| :||||| :||||| :|||||
228 ACCTTCATTTGATTCATGCAATGTGGCGGTACAGCATCTCTGATCCAC 277
   ||| :||||| :||||| :||||| :|||||
1960 uLeuProLeu.....CysGlnArgLysGluProLysGluAlaSerLysG 1975
   ||| :||||| :||||| :||||| :|||||
278 GTGCTATTTTGTACACACACTGTGTCATTTCGTTTCATTCATTATTT 327
   ||| :||||| :||||| :||||| :|||||
1975 InSerLeuAlaValThrThrLysAlaIleIle..... 1985
   ||| :||||| :||||| :||||| :|||||
328 GTTACAAAGTTGATGCGATATCGATACGATACGTTTACATGGAAGC 377
   ||| :||||| :||||| :||||| :|||||
1986 ...ThrHisValGluAsnThrSerArgIleTyrLeuGlnPheSerGly 2001
   ||| :||||| :||||| :||||| :|||||
378 TGAT..... 381
   ||| :||||| :||||| :||||| :|||||
2001 saSpSerLeuMetAspIleIleCysGluLysLeuAsnGlySerLysLeuG 2018
   ||| :||||| :||||| :||||| :|||||
382 .....AAACAGTTAGTCACACATTCAG..... 405
   ||| :||||| :||||| :||||| :|||||
2018 InProLysThrGluLysAlaAlaValAspAspMetCysValValGlnPhe 2034
   ||| :||||| :||||| :||||| :|||||
406 .....GTATCTGAATCTCAACTGC 425
   ||| :||||| :||||| :||||| :|||||
2035 AlaAspAspLeuGluPheTyrArgSerArgIleLeuGlnValLeuGluAs 2051
   ||| :||||| :||||| :||||| :|||||
426 TTTTCAAACTCAAAATGTGCCGATGCCAGTATGCCGTTATGAATTTGG 475
   ||| :||||| :||||| :||||| :|||||
2051 paSpGlnTyrLysValIle.....LeuIle 2060
   ||| :||||| :||||| :||||| :|||||
476 ATGCTGACCAACC.....GCTACACCACTTCAA 504
   ||| :||||| :||||| :||||| :|||||
2060 spTyrGlyAsnThrThrValValAspLysLeuTyrGluLeuProGlnGlu 2076
   ||| :||||| :||||| :||||| :|||||
505 TTTGCTATCATGTGTCAGCAGTTTATCATAAATGACATCGATCTGCA 554
   ||| :||||| :||||| :||||| :|||||
2077 PheThrLeuIle...LysProValAlaGlu.....IleCysSerMetG 2090
   ||| :||||| :||||| :||||| :|||||
555 A.....ACCGTTGATCACTTCT 571
   ||| :||||| :||||| :||||| :|||||
2090 uProSerAlaIlePheGluLysAsnLysAlaLeuThrLeuThrPheA 2107
   ||| :||||| :||||| :||||| :|||||
572 GCGCGGTGTCATTCTCTGCTTTGTGATGATGCTAACGGTACTGTCG 621
   ||| :||||| :||||| :||||| :|||||
2107 spAlaLeuLeuAspSerCys..... 2113
   ||| :||||| :||||| :||||| :|||||
622 GAAATTCCTAAATGCTGATGATGCTGCTTGTGATAAATTTTGTAAATAA 671
   ||| :||||| :||||| :||||| :|||||
2114 .....LysGlyValValAlaValGluPheValAsnLysSe 2125
   ||| :||||| :||||| :||||| :|||||
672 TTTGGAATATCCAAAGATTTAATGCTGCGCAAGAAGCTCACGATATACA 721
   ||| :||||| :||||| :||||| :|||||
2125 rAlaSerProProValValArgLeuThrThrLysAspLysArgSerLeu 2142
   ||| :||||| :||||| :||||| :|||||
722 AATATGCGATGCATCAGAGCTTTTGTATCAATGCCAGATCAGTATTACC 771
   ||| :||||| :||||| :||||| :|||||

```

```

2142 ysiLeTyrGlnHisLeuGlnLysLeuValGlnAlaGluLeuLysLeuIle 2158
772 ATTTAA...GAACCAATAGCGCAATGTTCGACCAACATGTTACAGAAC 818
   ||| :||||| :||||| :||||| :|||||
2159 GlnLysArgAsnGlnLysAsnSerGlnCysIleIleSerTyrGlyAsnSerP 2175
   ||| :||||| :||||| :||||| :|||||
819 ACAAGATTTCGAGCTGTTAAACAGAGTGTCGCCGACGCAAAACCTGCTG 868
   ||| :||||| :||||| :||||| :|||||
2175 oLysSerPheTyrValGlnMetLysHisAsnSerAlaAspLeuAspLeuI 2192
   ||| :||||| :||||| :||||| :|||||
869 CAGCTGCGCACTTCGTTACTACAGAAAAGATCTGCACAGCCGGAGCAT 918
   ||| :||||| :||||| :||||| :|||||
2192 LeValLysThrLeuGlnSerLeuLysGlnLysLeu.....LysLys 2206
   ||| :||||| :||||| :||||| :|||||
919 ATCATGATGATACGACTGATATCAACACCCCTTGAATTAAGCATGATA 968
   ||| :||||| :||||| :||||| :|||||
2207 LeuIleAspProThrThrAsnSerAsnGlyValCysTyrSerGlnLys 2223
   ||| :||||| :||||| :||||| :|||||
969 T 969
2223 p 2223

seq_name: SwissProt_38:SSR5_RAT
seq_documentation_block:
ID SSR5_RAT STANDARD; PRT; 363 AA.
AC P30938;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE SOMATOSTATIN RECEPTOR TYPE 5 (SSSR).
CN SSR5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PIUITARY;
RX MEDLINE: 93125499.
RA O'CARROLL A.-M., LOCAIT S.J., KONIG M., MAHAN L.C.;
RT "Molecular cloning and expression of a pituitary somatostatin
  receptor with preferential affinity for somatostatin-28.";
RL Mol. Pharmacol. 42:939-946(1992).
RN [2]
RP REVISIONS TO C-TERMINUS.
RC TISSUE=PIUITARY;
RX MEDLINE: 94195267.
RA PANETTA R., GREENWOOD M.T., WARSZYNSKA A., DEMCHISHYN L.L., DAY R.,
  NIZNIK H.B., SRIKANT C.B., PADEL Y.C.;
RT "Molecular cloning, functional characterization, and chromosomal
  localization of a human somatostatin receptor (somatostatin receptor
  type 5) with preferential affinity for somatostatin-28.";
RL Mol. Pharmacol. 45:417-427(1994).
CC -|- FUNCTION: RECEPTOR FOR SOMATOSTATIN-28. THE ACTIVITY OF THIS
  RECEPTOR IS MEDIATED BY G PROTEINS WHICH INHIBIT ADENYLYL CYCLASE.
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -|- TISSUE SPECIFICITY: PROMINENT IN THE PITUITARY AND SMALL
  INTESTINE. LOW LEVELS IN ISLETS AND SPLEEN. NOT DETECTED IN
  KIDNEY, PANCREAS, CEREBELLUM, OR CORTEX.
CC -|- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: I04535; AAA17029.1; -.
CC EMBL: U01152; AAC09011.1; -.
CC EMBL: X74828; CAA52825.1; -.
CC GCRDB: GCR_0514; -.

```


Ratio: 0.592 Gaps: 13
Percent Similarity: 45.652 Percent Identity: 18.944

alignment_block:

US-09-323-427-3 x AMD_HUMAN ..

Align seg 1/1 to: AMD_HUMAN from: 1 to: 974

```

97 TTGGACCACTTCATTAACATCATTTTAAATACAGTAATGATTCGA 146
   ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:
631 CysGlnProThrAspValAlaValAspProGlyThr..... 642
147 AGGACATGTTTATGTGAAGGCTTTATGTATCAAGAAGTTGCCGAATG 196
   ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
643 .GlyAlaIleTyrValSerAspGlyTyr..... 651
197 ATGAAGTGTGACGTCAAGTTGCCGAATTCCTCATTCATTCATTCG 246
   ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
652 ..... 652
247 AATGTTGGCGGTACAGCATCTGTGAATCCAGTGGTATTTTGTGAACAC 296
   ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
653 Asn...SerArgIleValGlnPheSerProSerGlyLysPheIleThrGI 668
297 A..... 301
668 nTPrGlyGlnGluSerSerGlySerSerProLeuProGlyGlnPheThrV 685
302 TT.....GTCAATTGCTTTCATCCATTATT.....GTTACCAAAATT 339
   ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
685 alProHisSerIleuAlaIleuValProLeuLeuGlnLeuCysValAla 701
340 GATCGTCGATAT...CGATGACAATGCTTTTACATGGAAGCTGATAAAC 386
   ||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
702 AspArgGluAsnGlyArgIleGlnCysPheLysThrAspThrLysGluPh 718
387 AGTTAGTGCACAGATTGAGGTATCTGAATCACAACACTGTTTCAAACTC 436
   ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
718 eValArgGluIleLysHisSerSerPheGlyArgAsnValPheAlaIle 725
437 AATTGTCCCGATGCCAGTATGCCGTTATGCAATTTTGGATGGTGACCA 466
   ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
735 eArgYrIlePro..... 738
487 ACCGGTCAACCAAGTTCATTTGCTATCATGTGTCAGCCAGTTATGATAA 536
   ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
739 .....GlyLeuLeuPheAlaValAsnGlyLysProHisPheGly.. 751
537 ATGGACATCGGATTTGAAACCTTGATCTTCTGCGCGGTTGCCATT 586
   ||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
752 .....AspGlnGluProValGlnGlyPhe..... 759
587 CCTGCTTTGCGATGATGTAACGGTGATCTGTGGAATTTCAATGCT 636
   ||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
760 .....ValMetAsnPheSerAsnGlyLysIleLeuAspIlePhe..... 772
637 GATGATGTGCTCTGTGATTAATATTTGCTAAATATTTGGAATTCACAC 686
   ||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
773 .....LysProValArgLysHisPheAspMetProH 783
687 AGATTAAATGCTGGCCAGAACCTCAGTATACAAATATGCGGATCGAT 736
   ||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
783 sAspIleValAlaIleSerGlnAspGlyThrValYrIleGlyAspAlaHis 800
737 CACAGCTTTTCTATCAATGCCAGATCAGTATTCACATTAAGAACCAAT 786
   ||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
800 hAsnThrValTyrLysPheThrLeuThrGluLysLeuGlnHisArgSer 816
787 AGCGAATGTGTTGACCAACAATGTTGACAAGCAAGATTCGAGACTGT 836
   ||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
817 ValLysLysAlaIleGlyIleGluValGlnGluIleLysGluAlaIle 833
837 TAAACAGGTGTGTGCGCAGCAAAACCTGTGTGACGTGCGCAACTTCGT 886

```

```

833 lValGluThrLysMetGlnAsnLysProThrSerSerGluLeuGlnLysM 850
887 TACTCAAGAAAGATCT.....GCAGACCGGAGATATCATGATGTA 930
   ||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
850 eGlnGluThrLysGlnLysLeuIleLysGlnProGlySerGlyValProVal 866
931 CGAAGTGAATTCACACCCCTTGAATTAAGCATGAATCAACCTTTGCC 980
   ||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
867 .ValLeuIleThrThrLeuLeuValIleProValValLeuLeuAlaI 883
981 AGTTGATTTTACGTC 994
883 lValIlePheIle 887
seq_name: SwissProt_38:CGEL_CHICK

```

```

seq_documentation_block:
ID CGEL_CHICK STANDARD: PRT: 407 AA.
AC P49707; 091032;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE G1/S-SPECIFIC CYCLIN E1.
GN CCNE1 OR CCNE.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RA LI H., LAHTI J.M., VALENTINE M., HOUSTON J., KIDD V.J.;
RL Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G1/S
CC (START) TRANSITION. INTERACTS WITH A MEMBER OF THE CDC2/CDK
CC PROTEIN KINASES FAMILY TO FORM MPF (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN E SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U28981; AAA74981.1; -.
DR EMBL: U28990; AAA81647.1; -.
DR EMBL: U28982; AAA81647.1; JOINED.
DR EMBL: U28983; AAA81647.1; JOINED.
DR EMBL: U28984; AAA81647.1; JOINED.
DR EMBL: U28985; AAA81647.1; JOINED.
DR EMBL: U28986; AAA81647.1; JOINED.
DR EMBL: U28987; AAA81647.1; JOINED.
DR EMBL: U28988; AAA81647.1; JOINED.
DR EMBL: U28989; AAA81647.1; JOINED.
DR HSSP: P30274; 1VIN.
DR PROSITE: PS00292; CYCLINS; 1.
DR PFAM: PF00134; cyclin; 1.
KW Cyclin; Cell cycle; Cell division.
FT CONFLICT 106 106 G -> GD (IN AAA81647).
FT CONFLICT 343 343 L -> S (IN AAA81647).
SQ SEQUENCE 407 AA; 46739 MM; 5664B6AB CRC32;

```

alignment_scores:

Quality: 88.50 Length: 313
Ratio: 0.637 Gaps: 16
Percent Similarity: 44.409 Percent Identity: 21.406

alignment_block:

US-09-323-427-3 x CGEL_CHICK ..

Align seg 1/1 to: CGEL_CHICK from: 1 to: 407

```

235 TTGTGATTCGCAATGTTGCGGTACAGATCT..... 267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
89 PheIleHisLeuAsnValAlaSerThrArgSerSerProLeuProIleLe 105
268 .....CTGAATCCACGTC 280
    ::||| ::
105 uGlyTrpAlaAsnArgAspValTrpIlyAsnMetIleAsnIlyGluG 122
281 GATATTTTGAACAACAAGCTGTGTCATTCGTTTCATCATCATATTT... 327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
122 LuThrTyValArgAspIlySerIlyMetGlnArgHisProLeuLeuGln 138
328 .....GTACCAAGT 338
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
139 ProIlyMetArgThrIleLeuLeuAspTrpLeuMetGluValGlyGluVal 155
339 TGCATCGTCATATCGAGTACATGCTTTTACATGAGAGCTGAT..... 381
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155 LTyTrIlySerIlyArg...GluThrPheTyLeuAlaGlnAspPheA 171
382 .....AAACAGTTAGTCACAG 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
171 spArgPheMetAlaThrGlnGlnAsnValValIlyThrLeuLeuGlnLeu 187
400 ATTGAGGATTCGAAATCACAACCTGCTTTTCAA...ACTCAATTTGCC 446
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
188 IleGlyIleSerSerIlePheIleAlaIlySerIlyGluIlyLeuTrp 204
447 GATCCAGATGCGCGTTATGAA...ATTGTGATGCTGACACACCGGTC 493
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
204 ProIlyLeuHisGlnPheAlaIlyThrAspGlyAlaCysThr.Glu 220
494 AACCAAGTTCATTTGCTATCATGTCACAGAGTTATCATTAATGAGACA 543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
221 AspGlnIleLeuSerMetGluLeuIleIleMetIlyAlaLeuAsnTrpAs 237
544 TGGCATTCGAAACCGTGATA.....CTTCTGCGCGGTTG 581
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
237 n.....LeuAsnProLeuThrValAlaSerTrpLeuAsnIleTyMetG 252
582 CCATTCCTGCTTTGTCATGATGTCAGCGTACTGTC..... 621
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
252 IValAlaIlyTrpLeuAsnGlnIlyLeuTyGluValLeuLeuProGlnTyPro 268
622 .....GAATTCCTAAATGCTGATGATGATGTC 647
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
269 GlnGlnIlePheValGlnIleAlaGlnLeuLeu.....AspLeuCysVal 283
648 TCTGTAAATATTTTCTAAATATTTTGAATATCCACAGATTTTAATG 697
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
283 IleuAsp.....IleGlyCysLeuGlnIlyTrpTyGlyValLeuA 297
698 CTGGCCAAAGAGCTCAGTATACAAATATGCGGATGATCAGAGCTT... 744
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
297 IAlaSerIleLeuTyHisPheSerSerSerGlnLeuMetGlnIlyVal 313
745 .....TTGTATCAATGCSAGATCATATTACATTAAGAACAATATG 788
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
314 SerGlyTyGluTrpCysGluLeu.....G 322
789 CGAATGTTTGCACACAAATGTTGACAAACCAAGATTCGAGATGTTA 838
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
322 uGlyCysValIlyTrpMetValProPheAlaMet.....AlaIleA 336
839 AAACAGAGTGGTCCGACGACAAAACCTGCTGCAGCTCGACAACTTCGTTA 888
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
336 rGlyIlyValAlaIlySerSer.....LysLeu 343
889 CTCGAAGAAAAGATCTGCAGAACCGGAGAAATATGATGATGATCAGACTGA 938
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
344 LysHisPheArgGlyIleAlaProGlnAspLeuHisAsnIleGlnThrHis 360

```

```

939 TATCAACACCCCTGAATTAAGCATGATATCAAGCT 975
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
360 sIleAsnSerIleAspLeuLeuAsnIlyValGlnAla 372

seq_name: SwissProt_38:PHY3_AVEsa
seq_documentation_block:
ID PHY3_AVEsa STANDARD: PRT; 1128 AA.
AC P06593.
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PHYTOCHROME A TYPE 3 (AP3).
GN PHYA3 OR PHY3.
OS Avena sativa (Oat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC Poaceae; Avena.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 86093660.
RA HERSHEY H.P., BARKER R.F., IDLER K.B., LISSEMORE J.L., QUAIL P.H.;
RT "Analysis of cloned cDNA and genomic sequences for phytochrome:
RT complete amino acid sequences for two gene products expressed in
RT etiolated Avena."
RL Nucleic Acids Res. 13:8543-8559(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 88185837.
RA HERSHEY H.P., BARKER R.F., IDLER K.B., MURRAY M.G., QUAIL P.H.;
RT "Nucleotide sequence and characterization of a gene encoding the
RT phytochrome polypeptide from Avena."
RL Gene 61:339-348(1987).
RN [3]
RP SEQUENCE OF 1-12.
RA GRIMM R., KELLERMANN J., SCHAEFER W., RUEDIGER W.;
RT "The amino-terminal structure of oat phytochrome."
RL FEBS Lett. 234:497-499(1988).
CC -1- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT
CC ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS
CC MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT
CC ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCOVERSION OF THOSE
CC PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS
CC RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE
CC RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR
CC GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-
CC BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,
CC PROTOCHLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS
CC THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.
CC -1- SUBUNIT: HOMODIMER.
CC -1- PFM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.
CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X03242; CAA26999.1; -.
DR EMBL: M18822; AAT76820.1; -.
DR PIR: A29631; A29631.
DR PIR: S00096; S00096.
DR PIR: S03032; S03032.
DR MENDEL: 1305; AVEsa; PHYA.1.
DR PROSITE: PS00245; PHYTOCHROME_1; 1.
DR PROSITE: PS50046; PHYTOCHROME_2; 1.
DR PFAM: PF00360; phytochrome; 1.
DR PFAM: PF00512; signal; 1.
DR PFAM: PF00989; PAS; 2.
DR PFAM: PF01590; GAF; 1.

```

KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;
 KW Multigene family.
 FT INIT MET 0 0
 FT BINDING 321 321 CHROMOPHORE
 FT CONFLICT 278 278 L -> F (IN REF. 2).
 SQ SEQUENCE 1128 AA; 124861 MW; 49DNA285 CRC32;

alignment_scores:
 Quality: 86.50 Length: 286
 Ratio: 0.618 Gaps: 21
 Percent Similarity: 48.951 Percent Identity: 23.776

alignment_block:

US-09-323-427-3 x PHY3_AVES

Align seg 1/1 to: PHY3_AVES from: 1 to: 1128

```

471 TTGGAGGTGGACCAACCGGTCAACGTTCAATTTGCTATCATTTGTC 520
    |||||
771 PheGlyTrp.....CysSerGluTrpAs 778
521 AGCCAGTTTATCATAA.....TGACATGCCATTTGAAACCGTT 561
    :|||:
778 nAlaAlaMetThr.LysLeuThrGlyTrpAsnArgasp..GluValLeu 793
562 GATACCTTCTCGCGGTTGTCATTCCTGCTTGTGATGATGATACGG 611
    ||| :
794 AspLysMet.....LeuLeuGlyGluValPhe.....AspSerSerAs 806
612 TGATCTGTGGAATTTCTAAATGCTGATGATGCTCTTGTATAAATAT 661
    : :
806 nAlaSerCysProLeuLysAsnArgAspAlaPheValSerLeuCysVal 823
662 TCGTAATATTTTGGATATCCACACAGATTATGGCTGGCCAAGAGCT 711
    |||:
823 euLeuAsnSer.....AlaLeuAlaGlyGluGluThr 833
712 CAC.....GTATCAAAATATGCGGATCGATCGACGTTTCTATCATG 755
    : :
834 GluLysAlaLysProPheGlyPhePheAspArgSerGlyLysTrpIleGly 850
756 CCAGATCAGTATTACCATTAAGAACCAATAGCGAA..... 792
    : :
850 sLeuLeuSerAlaAsnArgLysGluAsnGlyGlyLeuIleThrGly 867
793 ....TGTGTGACCAATGTTGACAGAACCAAGATTCGAGCGCTG 837
    ||| |||
867 alPheCys.PheIleHis..ValAlaSerHisGluLeuGlnHisAlaLe 882
838 A.....AAACAGGTGTGCGCGACGAAACCTGTCAGCTCGCAACT 881
    | :
882 uGlnValGlnGlnAlaSerGluGlnTrpSerLeuLysArgLeuLysAla 899
882 TCGTTTACTCAAGAAAGATCTCAGAACCGGAGAAATATGATGATGAC 931
    || |||
899 heSerTyMetArgHisAlaIleAsnAsnProLeuSerGlyMetLeuTy 915
932 GA.....ACTGATATCAAC.....AC 947
    |||||
916 SerArgLysAlaLeuLysAsnTrpAspLeuAsnGluGlnMetLysGln 932
948 CATTGAATTAAGCATGAT.....AATCAAGCTTTGCCAG 982
    : :
932 nIleHisValGlyAspAsnGlyHisGlnIleAsnLysIleLeu...A 948
983 TTGATTTTACGTACCGGTGACACTTCTGCACATATATGA..... 1020
    : :
948 lAspLeuAspGlnAspSerIleThrGluLysSerSerCysLeuAspLeu 964
1021 .....CAACCTGTAATACTGCTGCAGTACAA... 1047
965 GluMetAlaGluPheLeuLeuGlnAspValValAlaAlaValAlaSerGln 981

```

```

1048 .....AATGGAATC.....TGCATGNCAC 1066
    : :
981 nValLeuIleThrCysGlnGlyLysGlyIleArgIleSerCysAsnLeuP 998
1067 CATTGGCTTCTCAATGTTTATG.....GGTTTA 1095
    || :
998 ro.....GluArgPheMetLysGlnSerValTyrgLysAspGlyVal 1011
1096 AGCATGCAATGATGCTGCCGTCATATTACCATTTGCTTAATTTGC 1145
    : :
1012 ArgLeuGlnGlnIleLeuSerAspPheLeuPheIleSerValLysPhe 1028
1146 TCCA 1149
    |||
1028 rPro 1029

```


THIS PAGE BLANK (USPTO)

OM of: US-09-323-427-3 to: SPTREMBL_12.* out_format : pfs

Date: Apr 16, 2000 4:56 AM

About: Results were produced by the Gencore software, version 4.5.
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=framed+np2.model -DEV=xlp  
-O/cgcn2.1/USFTO.spool/US09323427/runat_14042000_170514_19920/app_query.fasta.1  
-DB=SPTREMBL_12 -OFMT=fastan -SUFFIX=backtrains.rspt  
-GAPOP=12.000 -GAPEXT=4.000 -MISMATCH=0.100 -LOOPEXT=0.000  
-LOOPEXT=0.000 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000  
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000  
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1  
-MATRIX=blomsim62 -TRANS=human40.cdi -LIST=45 -DLOCALIGN=200  
-THR_SCORE=pct -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORR=ext  
-MINLEN=0 -MAXLEN=100000 -USER=US09323427 -NCPU=6 -ICPU=3  
-NO_XLPHY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-323-427-3  
Query length: 1161  
Database: SPTREMBL_12.*  
Database sequences: 225878  
Database length: 69334122  
Search time (sec): 253.700000
```

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
sp_invertebrate:093115	1653.50	2830.20	2.8e-150	385	093115 ascaris lumbricoide
sp_invertebrate:019707	1382.50	2363.51	2.7e-124	389	019707 caenorhabditis elegans
sp_invertebrate:018479	1313.50	2243.77	1.2e-117	433	018479 meloidogyne artiella
sp_invertebrate:093409	725.50	1232.91	2.8e-61	364	093409 caenorhabditis elegans
sp_invertebrate:093775	695.00	1185.43	2.2e-58	206	093775 wuchereria bancrofti
sp_invertebrate:021808	652.00	1132.89	2.9e-55	127	093775 brugia malayi
sp_invertebrate:021540	654.50	1112.62	1.7e-54	290	021808 caenorhabditis elegans
sp_invertebrate:021540	602.50	1017.49	1.8e-49	550	021540 caenorhabditis elegans
sp_invertebrate:019053	595.00	1005.71	9.4e-49	484	019053 caenorhabditis elegans
sp_invertebrate:093532	578.50	981.09	3.4e-47	315	093532 caenorhabditis elegans
sp_invertebrate:022164	546.50	924.25	4.1e-44	384	022164 caenorhabditis elegans
sp_invertebrate:093409	511.00	859.04	1.1e-40	610	093409 caenorhabditis elegans
sp_invertebrate:023097	477.50	804.24	1.7e-37	440	093967 caenorhabditis elegans
sp_invertebrate:023097	476.00	802.82	2.4e-37	387	023097 caenorhabditis elegans
sp_invertebrate:022680	453.50	760.05	3.6e-35	609	022680 caenorhabditis elegans
sp_invertebrate:018213	414.00	696.16	2.1e-31	382	018213 caenorhabditis elegans
sp_invertebrate:009586	409.50	669.67	6.1e-31	747	009586 caenorhabditis elegans
sp_invertebrate:021573	401.50	620.43	3.5e-30	670	021573 caenorhabditis elegans
sp_invertebrate:021573	361.00	597.70	5.5e-26	66	096776 brugia pahangi
sp_invertebrate:019304	357.50	571.71	3.0e-24	437	019304 caenorhabditis elegans
sp_invertebrate:022627	293.00	488.21	8.3e-20	366	022627 caenorhabditis elegans
sp_invertebrate:045890	250.50	484.45	1.4e-19	344	045890 caenorhabditis elegans
sp_invertebrate:016304	158.00	252.77	7.7e-07	514	016304 caenorhabditis elegans
sp_invertebrate:023587	141.00	206.53	4.2e-05	3507	023587 caenorhabditis elegans
sp_invertebrate:018117	137.50	209.54	8.0e-05	1262	018117 caenorhabditis elegans
sp_invertebrate:022724	132.50	206.59	0.0002	665	022724 caenorhabditis elegans
sp_invertebrate:024328	126.50	195.85	0.0008	697	024328 drosophila melanogaster
sp_invertebrate:001494	120.00	181.37	0.0037	1011	001494 caenorhabditis elegans
sp_invertebrate:020167	116.00	177.11	0.0086	751	020167 caenorhabditis elegans
sp_invertebrate:018298	110.50	165.64	0.0299	942	018298 caenorhabditis elegans
sp_invertebrate:097298	105.00	148.12	0.1138	2340	097298 plasmodium falciparum
sp_invertebrate:020550	103.50	154.72	0.1380	828	020550 caenorhabditis elegans
sp_mammal:062672	101.00	144.66	0.2615	1589	062672 bos taurus
sp_vertibrate:091697	100.00	147.62	0.3038	935	091697 xenopus laevis
sp_human:005707	99.50	147.68	0.3346	843	005707 homo sapiens
sp_mammal:018758	98.00	120.72	0.6735	13288	018758 sus scrofa domestica
sp_mammal:077509	95.00	135.95	0.9598	1323	077509 bos taurus
sp_bacteria:097EG6	94.00	139.55	1.10	724	097EG6 carobacterium diversus
sp_bacteria:051229	93.50	140.21	1.21	610	051229 borrelia burgdorferi
sp_invertebrate:097239	93.00	122.87	1.73	3933	097239 plasmodium falciparum

sp_vertibrate:097239 + 92.50 133.56 1.62 1065
sp_human:094829 + 91.50 132.73 1.99 963
sp_invertebrate:096440 - 90.50 141.49 2.12 294
sp_virus:066951 + 90.50 127.30 2.63 1464

seq_name: sp_invertebrate:093115

```
seq_documentation_block:  
ID 093115 PRELIMINARY; PRT: 385 AA.  
AC 093115.  
DT 01-FEB-1997 (TREMBLER, 02, Created)  
DT 01-FEB-1997 (TREMBLER, 02, Last sequence update)  
DT 01-NOV-1998 (TREMBLER, 08, Last annotation update)  
DE CUT-1-LIKE CUTICLIN PROTEIN PRECURSOR.  
GN ASCUT-1.  
OS Ascaris lumbricoide (common roundworm)  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Ascaridida;  
OC Ascarididae; Ascarididae; Ascaris.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 97390131.  
RA TIMONONI M., BAZZICALUPO P.,  
RT "cut-1-like genes of Ascaris lumbricoide";  
RL Gene 193:81-87(1997).  
DR EMBL: U73005; AAB6646.1; -  
KW Signal.  
FT SIGNAL.  
FT CHAIN 1 16 POTENTIAL.  
FT 17 385 CUT-1-LIKE CUTICLIN PROTEIN.  
SQ SEQUENCE 385 AA; 42408 MW; BDB1169 CRC32;
```

alignment_scores:

Quality: 1653.50 Length: 388
Ratio: 4.593 Gaps: 4
Percent Similarity: 92.784 Percent Identity: 80.155

alignment_block:

US-09-323-427-3 x 093115
Align seg 1/1 to: 093115 from: 1 to: 385

4	ATGATTCGCTTATTCGCTTCTGACTACACTTATTCGCTTATTC	53
1	MetCysArgAlaValSerPhe...LeuAlaIlePheGlyLeuAlaIleAla	16
54	GATTCGGTTCGACATGCTGTCGACGTCGACGACGACGACGACGAC	103
16	alleProValAspAsnGlyValGlnGlyIleupProGlnIleGlnCysGlyP	33
104	CAACTTCATTAACAAATCAATTTTAATACAGTATATGATTCGACGACAT	153
33	rothserIlethrValAsnPhenAsnThrArgAsnProthegIleGlnHis	49
154	GTTTATGTGAAGGCTTTATGATCAAGAAGGTTGCCGTAATGATGATGAC	203
50	AlaTyrValIleGlyLeuTyrAspGlnGlnGlnCysArgSerAspIleu	66
204	TGACGTCGACGTCGACGTCGACGTCGACGTCGACGTCGACGTCGAC	253
66	YelYArgGlnValAlaGlyIleSerLeuProPheAspSerCysAsnValA	83
254	CGCGTCACGACGTCGATTCGACGTCGATTCGATTCGATTCGATTCGATTC	303
83	IaArgThrArgSerLeuAsnProArgGlyIlePheValThrThrVal	99
304	GTCATTCGCTTTCATTCATTTATTTATTCACCAAGTTCGATTCGATTCG	353
100	ValIleSerPheHisProleuPheIleThrValAspArgAlaTyrArg	116
354	AGACATGCTTTTACATGGAAGCTGATTAACAGTTCGATTCGACGATTCG	403
116	GValGlnCysPheTyrMetGlnAlaAspLysThrValSerThrGlnIleG	133

```

404 AGGATCTGAAATCACAACACTGTTTCAAACTCAAAATGTCCGATGCCA 453
133 LuValSerGluIleThrAlaPheGlnThrGlnIleValaPrometPro 149
454 GTATGCCGTTATGAATTTGGATGGGACCAACGGGTCAACCAAGTTCA 503
150 ValCysArgTyrGluIleLeuAspGlyGlyProThrGlyGlnProIleG 166
504 ATTGCGATGATGGTCAGCCAGCTTATCATTAATGACATGGCATCTG 553
166 nPheAlaThrIleGlyGlnGlnIleValaTyrHisLysTyrThrCysAspSerG 183
554 AAACCGTTGATCTTCTCGCGCGTGTCCATTCCTGCTTGTGATGAT 603
183 LuThrValaSerThrPheCysAlaValaValaHisSerCysPheValaAsp 159
604 GGAACCGGTGATCTGTGAAATTTCTAAAGCTGATGATGCTGCTCTGA 653
200 GlySerGlyAspThrIleGlnIleLeuAsnGlyGlnGlyCysAlaLeuAs 216
654 TAAATATTTCTAAATATTTGGAATTCACACAGATTTAATGGCTGGCC 703
216 pLysTyrLeuLeuAsnAsnLeuGlyTyrProThrAspLeuMetAlaGlyG 233
704 AAGAAGCTCAGTATACAAATATGCGGATGATCACAGCTTTCTATCAA 753
233 InGluAlaHisValaTyrLysTyrAlaAspArgSerGlnLeuPheTyrGln 249
754 TGCGACATGCTATTTACCATTAAGAACCAATAGCAATGTGTGACCC 803
250 CysGlnIleThrIleThrIleLysGlnProAsnSerGlyCysProArgPr 266
804 ACATGTTTCAACACCAAGATTCGAGCTGTTAAACAGGTGTGCC. 852
266 oThrCysSerGlnProGlnGlyPheGlyAlaValaArgProGlyGlySer 283
853 ..GCAGCAAACTGCTGCAGCTGCCCACTGCTTACTCAAGAAAGA 900
283 LeuAlaProLysLysGlnArgArgGlyGlnLeuArgLeuIleLysSer 299
901 TCGCAAGAACCGGAGATATCATTTGATGATGACAGATATCAACCCCT 950
300 GlyGlyAspTyrAspAsnThrLeuAspValaArgThrAspPheSerAlaLe 316
951 TGAATTAAGGATGATTAATCAAGCTTTGCCAGTTGATTTACGTACCCGTG 1000
316 uAspIleSerAspArgAspGlnAlaLeuPrometAspLeuArgHisArg 333
1001 CACTTCTGCACAT...ATGACAACTGTAACTTCTGTCGACATACAA 1047
333 La.....ArgHisAlaArgGlyGlnGlnIleLeuSerProAlaAsn 347
1048 AATGSAATCGCATGTCACCATTTGGTCTCAATGTTTATGGGTTAAG 1097
348 GlnGlyIleLeuSerProPheGlyPheSerIlePheMetGlyLeuAl 364
1098 CATTGATGATGCTGCCGTCATTTATACCATTTGCTTAAATTTGCTG 1147
364 aValaAlaLeuAlaAlaAlaValaValaValaValaSerPheLysLeuArgP 381
1148 CAATCAGAAGCA 1161
381 roGlnGlnLysAla 385
seq_name: sp_invertebrate:Q19707

```

```

GN F22B5.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA Sims M.;
RL Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSWORTH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPELY T., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL: Z50044; CAA90355.1;
SQ SEQUENCE 389 AA; 42983 MW; 56B9543C CRC32;

```

```

alignment_scores:
  Quality: 1382.50      Length: 385
  Ratio: 4.177          Gaps: 8
  Percent Similarity: 85.974  Percent Identity: 69.610

```

alignment_block:

US-09-323-427-3 x Q19707 ..

Align seg 1/1 to: Q19707 from: 1 to: 389

```

34 CTATTTGATGCTGTTATTTGATTCGCGTTGACAAATGGTCCGAAGTGA 83
13 LeuValAlaSerValSerAlaIleProValaAspAsnValGlnGlyG 29
84 GCCAGAAATTTGAATGTGGACCAACTTCAATACATCAATTTTAATACAC 133
29 uProGlnValGlnGlyGlyProThrSerIleThrValaAspPheAsnThr 46
134 GTAAATGATTCGAAGACATGTTTATGTGAAGCTTTTATGATGTCAGAA 183
46 rGAsnAlaPheGlnGlnIleHisValaTyrValaLysGlyLeuPheAspGlnG 62
184 GGTGGCGTAAATGATGAAGGTGAGCTCAAGTTGCCGGAATTTCACTTGC 233
63 GlnCysArgAsnAspArgGlnGlyArgGlnValaAlaGlyIleGlyLeuP 79
234 ATTTGATTCATGCAATGTTGCCGCTACACGATCTGAATCCACGTGTA 283
79 oPheAspThrCysAsnValaAlaArgThrArgSerLeuAsnProGlyGly 96
284 TTTTGTGTACAAACAACACTGTTGTCATTTGCTTTCATCATTTATTTGACC 333
96 alPheValThrThrThrValaValaValaSerPheHisProGlnPheValThr 112
334 AAGTTGATCGTGCATATGAGTACAGTCAATGCTTTACATGGAAGCTGATA 383
113 LysValaAspArgAlaTyrArgValaGlnCysPheTyrMetGlnAlaAsp 129
384 AACATTAGTGCACAGATTGAGGTATCTGAATCACAACACTGCTTTTCAAA 433
129 sThrValSerThrGlnIleGlnValaSerAspLeuThrThrAlaPheGln 146
434 CTCAATTTGCTCCGATGCCAGTATGCGGTTATGAAATTTTGGATGGTGA 483
146 hrcGlnValaValaPrometProIleCysLysTyrGlnIleLeuAsnGlyGly 162

```

```

484 CCACCGGTCACACGATTCATATTCATTCATTCGTCACACGATTATCA 533
      |||||.....|||.....|||.....|||.....|||.....
163 ProthrglyluProvalglInphealThrilegylnglInglValtyrhi 179
      |||||.....|||.....|||.....|||.....|||.....
534 TTAATGACATCGCATTCGAAACCGTGATACCTTCTCGCGGTTGCC 583
      |||||.....|||.....|||.....|||.....|||.....
179 slsStrphrcysaspSerGluthrValasphrPhecysalavalValH 196
      |||||.....|||.....|||.....|||.....|||.....
584 ATTCCTGCTTGTGATGATGATGATGATGATGATGATGATGATGAT 633
      |||||.....|||.....|||.....|||.....|||.....
196 lSerCysThrValaspaspGlyasnGlyaspThrValglInlleuasp 212
      |||||.....|||.....|||.....|||.....|||.....
634 GCTGATGATGCTGCTTGTGATTAATATTTGCTAAATATTTGGAATATCC 683
      |||||.....|||.....|||.....|||.....|||.....
213 GluasnglyCysAlaleuaspPlyspPheleuaspAsnleuGlutyrPr 229
      |||||.....|||.....|||.....|||.....|||.....
684 AACAGATTTAATGCGTGGCCAGAAAGCTCAGATATACAAATGCGGATC 733
      |||||.....|||.....|||.....|||.....|||.....
229 oThrAspleuMetalaiglynglInglualahlsValtyrlyTyralaspA 246
      |||||.....|||.....|||.....|||.....|||.....
734 GATCAGAGCTTCTATCATGTCAGATCATATTCATTAAGAACCA 783
      |||||.....|||.....|||.....|||.....|||.....
246 rGserGlnleuPheThrglnCysGlnlleSerlleThrVallysgluPro 262
      |||||.....|||.....|||.....|||.....|||.....
784 AATPAGCAATGTGTTCGACACCAATGTTCCAGAACACAGATTCGAGAC 833
      |||||.....|||.....|||.....|||.....|||.....
263 AsnglInglCysAlaAlaGrProThrCysSerGluProglnglyPheglVal 279
      |||||.....|||.....|||.....|||.....|||.....
834 TGTTAACAGGTGTCGCCGACGAAACCTGCTGCAGCTGGCAGCACTTC 883
      |||||.....|||.....|||.....|||.....|||.....
279 aVallysglnlaaspGlnThrAlaGln.....Phephea 291
      |||||.....|||.....|||.....|||.....|||.....
884 GTTATCTCAGAAAAGATGTCAGAA...CCGAGAAATATCATTTGATGTA 930
      |||||.....|||.....|||.....|||.....|||.....
291 rValleuaspLysargSerAlaProvalMetgluAsnleuaspVal 307
      |||||.....|||.....|||.....|||.....|||.....
931 CGAAGTATATACACACCCCTTGAATATGCGATGATATCAAGCTTTGCC 980
      |||||.....|||.....|||.....|||.....|||.....
308 ArgAlaGlnleuThrThrleuGlnValleuGlnlyAsn.....LeuPr 322
      |||||.....|||.....|||.....|||.....|||.....
981 ACTGATTTAGCTCACCGCTGACTTCG.....CAACATATGAGAC 1021
      |||||.....|||.....|||.....|||.....|||.....
322 oSerSerleuThrGlnAlaGlnAlaValAlaSerArgGlnlleGlyG 339
      |||||.....|||.....|||.....|||.....|||.....
1022 AACCTGTATACTTGCTGCGATGACAAATGATGATGATGATGATGAT 1071
      |||||.....|||.....|||.....|||.....|||.....
339 luasp.....SerPheArgGlnleuLysleuSerSerPhe 351
      |||||.....|||.....|||.....|||.....|||.....
1072 GCGTCTCA.....ATGTTATAGGTTTAAAGCATTCGATTCATTCGC 1112
      |||||.....|||.....|||.....|||.....|||.....
352 HsllSerValValThrValPheleuGlyleuThrVal.....PheVa 366
      |||||.....|||.....|||.....|||.....|||.....
1113 TGCCGTCATTAATTAAC.....ATTTCGTTTAAATTTGTCGAATTC 1153
      |||||.....|||.....|||.....|||.....|||.....
366 lAlaIlePheIleThrTyMetlleValSerArgMetValProSerA 383
      |||||.....|||.....|||.....|||.....|||.....
1154 AGAAG 1158
      |||||.....|||.....|||.....|||.....|||.....
383 sPLys 384

```

seq_name: sp_invertebrate:018479

seq_documentation_block:

```

ID 018479 PRELIMINARY; PRT: 433 AA.
AC 018479;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DE 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE CUTICLIN-1.
GN mCUT-1.
OS Meloidogyne artiella.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Diplogasteria; Tylenchida;

```

```

OC Tylenchina; Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97218031.
RA DE GIORGI C., DE LUCA F., DI VITO M., LAMBERTI F.;
RT "Modulation of expression at the level of splicing of cut-1 RNA in the
RT Infective second-stage juvenile of the plant parasitic nematode
RT Meloidogyne artiella."
RL Mol. Gen. Genet. 253:589-598(1997).
DR EMBL: X96677; CAA65452.1; -.
KW Cuticle.
SO SEQUENCE 433 AA; 47582 MW; FDE26884 CRC32;

alignment_scores:
  Quality: 1313.50      Length: 418
  Ratio: 4.042          Gaps: 7
  Percent Similarity: 77.751      Percent Identity: 63.636

alignment_block:
  US-09-323-427-3 x 018479

Align seg 1/1 to: 018479 from: 1 to: 433

4 ATGATTCGTCCTTATTCCTTCTGTACTACACTTATTCATTCGCT.... 48
  |||.....|||.....|||.....|||.....|||.....
1 MetatgLyseuLeuPhealIalegIyValPheValAlaleuaspAlaI 17
  |||||.....|||.....|||.....|||.....|||.....
49 .....TATTCGATTCGCGTTGACAAATGCTGTCGAGGTGACCCAG 88
  |||||.....|||.....|||.....|||.....|||.....
17 ePheThrValAlaArgAlaIleProValaspaspGlyValGlyngluProG 34
  |||||.....|||.....|||.....|||.....|||.....
89 AATATGAATGATGGACAACTTCATATACATCAATTTATATACAGTAT 138
  |||||.....|||.....|||.....|||.....|||.....
34 lulleGlnCysgluProThrSerlleThrValaspPheAsnThrArgAsn 50
  |||||.....|||.....|||.....|||.....|||.....
139 CCATTCGAAGACATGTTTATGTTGAAGGCTTTATGATCAAGAAGTTG 188
  |||||.....|||.....|||.....|||.....|||.....
51 PropheGlnGlnHlsValtyrVallysglyleuPheaspGlnAlaGlyC 67
  |||||.....|||.....|||.....|||.....|||.....
189 CCGTATATGATGAAGTGGACGTCAGTTCGCGAATTCACCTTCATTTG 238
  |||||.....|||.....|||.....|||.....|||.....
67 sarArgSeraspGlnHlsGlyArgGlnValAlaGlylleGlnleuProphea 84
  |||||.....|||.....|||.....|||.....|||.....
239 ATTTCGCAATGTGCGGCTACACGATCTGGAATCCAGGTTATTTT 288
  |||||.....|||.....|||.....|||.....|||.....
84 sPserCysaspnValAlaArgThr...AspAlaGluProLysgluValPhe 99
  |||||.....|||.....|||.....|||.....|||.....
289 GTAACACACACTGTTGTCATTTGTTTCATTCATATTTGTTACCAAGT 338
  |||||.....|||.....|||.....|||.....|||.....
100 ValSerThrThrValValIleSerPheHlsProGlnPheValThrLysVa 116
  |||||.....|||.....|||.....|||.....|||.....
339 TGATGTCGATATTCGAGTACATGCTTTTACATGGAAGCTGATAAACG 388
  |||||.....|||.....|||.....|||.....|||.....
116 lAspArgAlaTyArgValGlnCysPheTyMetGlnAlaaspLysThrV 133
  |||||.....|||.....|||.....|||.....|||.....
389 TTAGTGCACAGATGAGTGTATGTAATCAACACAGCTTTTCAACTCA 438
  |||||.....|||.....|||.....|||.....|||.....
133 alSerAlaGlnleuGlnValSerGlnlleThrThrGlnPheGlnThrGln 149
  |||||.....|||.....|||.....|||.....|||.....
439 ATTGTCGCCGATGCCAGTATCCGTTATGAATTTTGGATGGGACCAAC 488
  |||||.....|||.....|||.....|||.....|||.....
150 ValValPrometProvalCysLysTyrgluIleleuGlnGlyAlaAlaIe 166
  |||||.....|||.....|||.....|||.....|||.....
489 CCGTACACCAAGTTCAATTTGCTATCATTCGTCAGCCAGTTTATCAAT 538
  |||||.....|||.....|||.....|||.....|||.....
166 uGlynglnProIleGlnPheAlaThrIleAlaGlnGlnValTyThrSerG 183
  |||||.....|||.....|||.....|||.....|||.....
539 GGACATGCGCATTCGTAACACCGTTGATCTTCTGCGCGGTTGCCATTC 588
  |||||.....|||.....|||.....|||.....|||.....
183 lYThrCysaspSerGluThrIleaspThrPhecysAlaValAlaHlsSer 199
  |||||.....|||.....|||.....|||.....|||.....
589 TCCTTTGTCGATGATGATGATGATGATGATGATGATGATGATGATGAT 638

```

```

11  |||
200 CysValValAspAspGlyAsnGlyAspThrValGlnLeuAsnGlu 216
639 TGGATGTGCTGATTAATATTCGTAATATTCGATATTCGACAC 688
216 ySerCysAlaLeuAspLysPheLeuLeuAsnLeuGluTyrProTha 233
689 ATTAAATGGCTGGCAAGAGCTACGATATCAATATGCGATCA 738
233 spruethrAlaGlyGlnGluAlaHisValTyrTyrAlaAspArgSer 249
739 CACCTTTTCATCATGCCGATGATATTCATTAACAACCAATAG 788
250 GlnLeuPheTyrGlnCysGlnIleSerIleThrIleGluProHisSe 266
789 CGAATGTGTGCGACCAATGTCACAGCAACCAAGATTGCGAGTCTTA 838
266 rGluCysAlaAlaTyrProLysCysAlaGluProSerGlyPheAsnAlaValL 283
839 AAACAGCT.....GATGCCGACAGCAAAACCTGCTGAGCTGCG 876
283 yValGlyAlaGlyAlaGlyAlaGlyAlaGlyAlaProProAlaAla 299
876 ..... 876
300 AlaAlaThrProProAlaAlaAlaProProAlaProIleAlaProAl 316
877 .....CAACTGCTTACTCA 892
316 alleProAlaAlaValAlaGlnArgLeuArgThrArgGlnLeuArgLeu 333
893 AGAAAGATCTGCAGAACG...GAGAAATATCATTTAGTACAGCTGAT 939
333 rGlySarGArgAlaAlaAlaProGlyGlnGlnValLeuAspValArgThrGln 349
940 ATCAACACCTTGAATATGACGATGATATCAAGCTTTGCGATGATT 989
350 PheAsnAlaLeuAspIleSerAspLysProPheAsnLeuProAlaHis 366
990 ACCTGACCGCTCTGTCACATTAAT.....GGACACCTGTATATAC 1033
366 eArgHisHisHisLeuMetCysAsnLeuSerAlaGluProGlnArgA 383
1034 TTGCTCAGTCAAAATGGA....ATGTCATGTCACCATTTGCGCTTC 1077
383 rPheIleTyrGlnGlyGlnGlnValCysMetSerSerPheArgAla 399
1078 TCAATGTTATGGC..TTTAAGCATGTCATTGATTCGCGCATTAATTA 1126
400 AsnHisLeuLeuArgTyrGlnHisCysAsnGlyCysSerHisTyrPyrA 416
1127 CCAT 1130
416 HIS 417
seq_name: sp_invertebrate:Q9XVM7
seq_documentation_block:
ID Q9XVM7 PRELIMINARY; PRT; 364 AA.
AC Q9XVM7.
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE F53P1.1 PROTEIN.
GN F53P1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE; 94150718.
RA BURTON J.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

```

```

RT elegans."
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,
RA CRAWTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATKINSON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL Nature 368:32-38(1994).
DR EMBL; 281088; CAB03124.1; -;
SQ SEQUENCE 364 AA; 40843 MW; 3FEC30BC CRC32;

```

```

alignment_scores:
  Quality: 725.50      Length: 271
  Ratio: 3.455         Gaps: 1
  Percent Similarity: 77.491  Percent Identity: 47.601

```

```

alignment_block:
US-09-323-427-3 x Q9XVM7

```

```

Align seg 1/1 to: Q9XVM7 from: 1 to: 364

```

```

55 ATTCGCGTTGACAAATGCTGTCGAGGTGAGCCAGAAATGAATGTCGACC 104
49 ValProIleGlnAsnSerLeuTyrGlyAspValGlnValGluCysAspSe 65
105 AACTCAATTAACATCAATTTTAATACAGCTAATGTCATTGCAAGCAGATG 154
65 rArgThrIleSerValGlnIleLeuThrGluLysProPheValGlyValI 82
155 TTTATGTGAAGGCTTTTATGATCAAGAGGTGCGCATGATGTAAGGT 204
82 IePheValLysAspPheAlaSerGluGluValCysThrSerArgGlyThr 98
205 GGACGTCAAGTTGCCGGAATTCACATTCGATTTGATTCGAAATGTCG 254
99 GlyArgLeuSerAlaPheLeuGluIleGluIleGlyLeuGlyAlaIle 115
255 GCGTACACGATCTCTGATCCAGCTGATTTTGTAAACAACACTGTG 304
111 ValGlnArgValLeuAsnProLysGlyLeuAlaValAlaArgThrIle 132
115 TCATTTGCTTTCATCATTAATTTGTTACCAAGTTGATGTCGATTCGA 354
305 hTrIleSerPheHisProTyrPheIleThrLysValAlaAspArgThrTyrAsn 148
132 hTrIleSerPheHisProTyrPheIleThrLysValAlaAspArgThrTyrAsn 148
149 LeuLeuCysLeuTyrArgLysSerGlnValThrValAlaAsnIleSe 165
405 GGTATGTGAATATCAACATCTTTTCAACCTCAAAATTTGCCGATGCCG 454
165 rValAspGluIleSerThrIleSerTyrAsnValAsnLeuThrMetPro 182
455 TATGCCGTTATGAATTTTGGATGGTGACCAACGCGTCACACGATTCGA 504
182 hrcysThrTyrGlnIleLeuSerGlyGlyProPheGlyLeuProValGlu 198
505 TTTGCTATCATTTGTCAGCCAGTTTATCAATAATGACATGCGATTCGA 554
199 PheGlyLeuIleGlyGlnGlnValTyrHisGlnTyrLysCysAspAsn 215
555 A.....ACGTTGATACCTTTCTGCGCGCTTGCTC 583

```

```

215 phyuglymetllelleleuilelysgluaspserserphecysmevalvalh 232
584 ATTCCGGCTTTGCGATGATGACGGATGATGATGATGATGATGATGATGAT 633
232 lstrhysservalaspaspglyarglyglthrserpheluleileasp 248
634 GCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 683
249 SerasncllyCysserileaspplyspheluleuSerasnleuclutyrtr 265
684 AACAGATTTAATGCGTGGCCCAAGACCTCAGTATACAAATATGCGATC 733
265 oglyasnleuleualaglyglngluahisvaltyrlysphealaaspa 282
734 GATACAGATTTTCTATCAATGCGATGATGATGATGATGATGATGATGAT 783
282 r9aspalaleuphepnehglnysglnlleserllefthrvallysgluapo 298
784 AATAGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 833
299 aspglnglucysvalarprollecgysgluaspsvalcluclygllyal 315
834 TGTAAACAGGT 846
315 aprovalvalgly 319

```

seq_name: sp_invertebrate:O9XYU9

seq_documentation_block:

```

ID O9XYU9 PRELIMINARY; PRT: 206 AA.
AC O9XYU9;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE CUTICULIN-1 (FRAGMENT).
OS Mucheria bancrofti.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;
OC Filarioidae; Onchocercidae; Mucheria.
RN [1]
RP SEQUENCE FROM N.A.
RA RAMZY R., HELMY H., ADELY M., CURTIS K., WEIL G.;
RT "Mucheria bancrofti L3 cuticulin-1 cDNA partial sequence.";
RL Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF125580; AAD28743.1; -.
FT NON-TER 1
SQ SEQUENCE 206 AA; 22602 MW; 4DBE20D0 CRC32;

```

alignment_scores:

```

Quality: 695.00 Length: 198
Ratio: 4.064 Gaps: 4
Percent Similarity: 86.364 Percent Identity: 69.192

```

alignment_block:

US-09-323-427-3 x O9XYU9 ..

Align seg 1/1 to: O9XYU9 from: 1 to: 206

```

556 ACCGTTGATCTTCTGCGGGTGTGCATTCCTGTTGTGATGATG 605
1 Thrvalasprthrphecysalavalahissercysphevalasprgi 17
606 TAACGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 655
17 yasnnglyasplyvalgluuleuasnlaasplycysalaleuasp 34
656 AATATTGCTAATATGATGATGATGATGATGATGATGATGATGATGATGATG 705
34 yspheuleuasnleuasnleuclutyrProthrspheuleuataaglyglh 50
706 GAAGTCACGTATACAAATATGCGGATGCATCAGCTTTTCTATCAATG 755

```

```

51 gluahisvaltyrlystyralaaspargselglnleuphetrglncy 67
756 CGAGTACATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 805
67 sglnlleserllefthrlysglnproasnserlycysgluarprog 84
806 AATGTCAGAACCCCAAGGATTCGAGCTGTAAACA...GGTGGTGC 852
84 Incystrhgurproglnglyphleglyalalelyshrhrgasnglyval 100
853 GCAGCAAAACCTGCTCAGCTCGCGCACTTGCTTACTCAAGAAAGATC 902
101 AlaalaalaserarglnlaalaalaphedrValleuclulsyars 117
903 TGCAGAACCGGAAATATGATGATGATGATGATGATGATGATGATGATGATG 952
117 pvalatrgaspgluasnlevalaspsvalargthrpsleuasalaleu 134
953 AATAGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 999
134 spleasnnglucyphethrserleuproasnlaaleuarhgsargsr 150
1000 GCACCTTCGCAACAT...AATGACAACTGTATATCTGCTGACGATCA 1046
151 Serleuleualahisgluasnglnhisprovallevalalathrmeth 167
1047 AATGGAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1083
167 rglnglyllecysmetsertlethrlyphethrleuualaglymetleu 184
1084 ..TTATGGTTAAGCATGATGATGATGATGATGATGATGATGATGATGAT 1125
184 lehevalallevalservalalathrlevalalathrleu 198

```

seq_name: sp_invertebrate:O96775

seq_documentation_block:

```

ID O96775 PRELIMINARY; PRT: 127 AA.
AC O96775;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)
DE CUTICULIN (FRAGMENT).
GN CUT-1.
OS Brugia malayi.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida;
OC Filarioidae; Onchocercidae; Brugia.
RN [1]
RP SEQUENCE FROM N.A.
RA LEWIS E.K., HUNTER S.J., TETLEY L., PAVIA NUNES C., BAZZICALUPO P.,
RA DEVANEY E.;
RT "cut-1-like genes are present in the filarial nematodes Brugia pahangi
RT and Brugia malayi and, as in other nematodes, code for components of
RT the cuticle."
RL Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AJ012617; CAA10074.1; -.
FT NON-TER 1
SQ SEQUENCE 127 AA; 14119 MW; 3E738071 CRC32;

```

alignment_scores:

```

Quality: 662.00 Length: 127
Ratio: 5.295 Gaps: 0
Percent Similarity: 98.425 Percent Identity: 96.063

```

alignment_block:

US-09-323-427-3 x O96775 ..

Align seg 1/1 to: O96775 from: 1 to: 127

```

304 GTCATTTGTTTCATCATTTATTTTACCAAGTGTGATGTCATATCG 353

```

1 ValIleSerPheHisProLeuSerValThrIlyValAspArgIaIyIar 17
354 AGCAACAAGCTTTTACATGGAAAGCTGATAAACAGTAAAGTCACACAGATTG 4030
17 gValGInCysPheTyIyMeGluAlaAspIlyShIyValSerThIeGInIleG 34
404 AGGATCTGGAAATCACAACTGCTTTTCAACAGCAAAATGTGCCGATGCCA 4533
34 IuValSerGluIleThrIAlaPheGlnIleGluIleValProMePto 50
454 GATAGCGGTATGAAATTTTGGATGGTGGACCAACCGGTCAACAGATTCA 5033
51 ValCysArgIyGluIleLeuAspGlyIyProThrIyGInProIleG 67
504 ATTTCGATCATTTGGTCGACCCAGTTATCATAAATGGACATGCATTTCTG 5533
67 nPheAlaThrIleGlyGInProValIlyHisIyStrPhnCysAspSerG 84
554 AAACCGGTGATCTTCTGCGGGTATGTCACATCCCTGTTTGTGGATGAT 6033
84 IuThrValAspThrIlePheCysAlaValIValHisSerCysPheValAspAsp 1000
604 GGTAACGGTGATCTGTGGAAATTTCAAATGCTGATGGATGCTGCTTGA 6533
101 GlYAsnGlyAspThrValGInIleLeuSerAlaAspGlyCysAlaLeuAs 1177
654 TAAATATTTCCTAAATATTTGGATATACCA 684
117 pIyStrIyLeuLeuAsnAsnIleGInIyPro 127

seq_name: sp_invertebrate:Q21808

```

seq_documentation_block:
ID      021808      PRELIMINARY;      PRT;      290 AA.
AC      021808;
DT      01-NOV-1996 (TReMBLrel. 01, Created)
DT      01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT      01-JAN-1999 (TReMBLrel. 09, Last annotation update)
DE      R07E3. 3 PROTEIN.
GN      R07E3. 3.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilitida;
OC      Rhabdilitina; Rhabdilitoidea; Rhabdilitidae; Pelodierinae; Caenorhabditis.
LN      [1]
RP      SEQUENCE FROM N.A.
RA      COTTAGE A.;
RL      Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RN      [2]
RX      MEDLINE; 94150718.
RA      WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA      BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA      CRAYTON M., DEAR S., DU Z., DUREIN R., FAVELLO A., FULTON L.,
RA      GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSON L.,
RA      JONES M., KERSHAU J., KIRSTEN J., LAISTER N., LATRILLE P.,
RA      LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA      PARSONS J., PERCY C., RITKEN J., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA      SMAULDON N., SMITH A., SONNHAMMER E., STADEN R., SULISTON J.,
RA      THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA      WATSON A., WEINSTOCK L., WILKINSON-SPOAT J., WOHLIDMAN P.;
RT      "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL      elegans.";
RL      Nature 368:32-38(1994).
RL      EMBL; Z49207; CAA89068.1; -.
SQ      SEQUENCE      290 AA; 33215 MW; 441587FE CRC32;

```

	alignment_scores:	
Quality:	654.50	Length: 283
Ratio:	3.306	Gaps: 8
Percent Similarity:	69.965	Percent Identity: 45.230

alignment_block:

```

US-09-323-427-3 x Q21808 ..
Align seg 1/1 to: Q21808 from: 1 to: 290

```

```

271 AACCCAGCGGCGATATTTTGTGTATCAACAACACTGTGTGATTCGTTTCAATCC 320
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4 SerProArg1LeuHeuSerThrsnValValAlaIleAsnPro 20
321 ATTATTTGTTACCAAGTTGATCGCGATATACGAGTATCAATGCTTTTACA 370
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
20 OGILHeuLeuThrLysAsnAspArgValPheLysValGlnProSyrHeuTyrM 37
371 TCGAAGCTGATTAACAAGTTACTAGCACAGATTGAGGTATCTGAATACAGA 420
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
37 etGluMetGluArg1IleGlnLysValIleGlnIleSerMetProPro 53
421 ACGCTTTTCAAACTCAAAATTTGCCGATGGCAAGATGCGCTTATGAAT 470
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
54 ProThrMetHisSerLysGlnLeuAsnMetProValCysLysTyrGluVal 70
471 TTTGGATGCGTGCACCAACCGGTCAACACGATTCATATTTGCTATCTTGTC 520
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
70 IleuAspGlySerProThrGlyProProValTyrPheAlaThrValGlyG 87
521 AGCCAGTTTATCATAATGGAACATGCGATCTCGAAGCCGTGATACTTTC 570
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
87 IlnMetValTyrHisLysTyrPheLysAspThrGlnHisGlnAspThrPhe 1030
571 TGGCGGTTGTCATTCCTCGCTTTTGCATGATGGTAAACGGTGTACTGT 620
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
104 CysMetLeuValHisSerCysPheValAspArgLysAsnGlyGlnArgVal 120
621 GGAATTTCTAATGCTGATGATGATGTGCTCTTGATTAATATTTGCTAAATA 670
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
120 IGlInLeuAsnAspLysGlyCysAlaLeuAspLysTyrLeuLeuThra 1370
671 ATTGTGAATTCACACAGTTTAATAGCTGGCGCAAGAGCTACAGTTTAC 720
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
137 snLeuGlnTyrProThrAspLysMetAlaGlyArgGluAlaHisValTyr 1530
721 AAATATGCGGATCGATCGACAGCTTTTCTATCAATGCGCATCACTATTAC 770
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
154 LysTyrAlaLysArgLysAspHisMetLysPheAspCysGlnIleSerIleTh 170
771 CATTAAGAACAACAATAGCGGA...TGCTGTTCGACACACATGTTCAGAAC 817
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
170 rValLysGlnProGlyLeuAsnProTyrCysAspValProSerCysProAsp 1870
818 CA..... 819
187 rOPrOPArgArgArgSerAsnThrLeuProAlaProAspAspAsnIle 2030
820 .....CAAGATTCGAGCTTTAAAC 842
204 ThrAlaIleAlaIleHisIleGlnTyrGluAspSerGlnIleLeuSerA 220
843 AG...GTGGTGGCGGACAGAAACCTGCTGACAGCTG.....CGC 877
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
220 sPryTyrIleIleProAsnAspAspIleIleSerLeuAsnThrPheGlnArg 236
878 AACCTGCGTTTACGCAAGAAAGATCTGCAGAAACGGGAATATCATTTGAT 927
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
237 AspAspHe.....AspMetArgIleSer..... 243
928 GTACGACACTGATATCAACACCCCTTGAAATTGACCATATATATCAAGCTTT 977
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
244 ....GluLeuCysMetThrAlaIleGlyThrThrLeuLeuValPheLeu 259
978 GCCAGTTGATTTACGTC.....ACGTCGACACTTCTGC 1009
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
359 snAlaPheLeuPheIleIleSerLeuValSerIleValHisValCys 274

```

seq_name: sp_invertebrate:Q21540

```

seq_documentation_block:
ID 021540 PRELIMINARY; PRT; 550 AA.
AC 021540:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE M142.2 PROTEIN.
GN M142.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA MCMURRAY A.;
RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,
RA THERRER-MEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
DR EMBL; 273428; CA97806.1;
DR PFAM; PF00092; vwa; 1
DR PRINTS; PR00453; VMPADOMAIN.
SQ SEQUENCE 550 AA; 61445 MW; A055FB45 CRC32;

alignment_scores:
Quality: 602.50 Length: 302
Ratio: 2.953 Gaps: 7
Percent Similarity: 67.550 Percent Identity: 43.377

alignment_block:
us-09-323-427-3 x 021540 ..
Align seg 1/1 to: 021540 from: 1 to: 550

79 GGTGAGCCAGAAATTGATGATGGACCACTTCATATCAATATCAATTTTAA 128
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
228 G1YthPrG1u1lellecysglProAspArg1leg1YallysAlase 244
129 TACACGTAATGATTCGAGACATGTTATGCGAAAGCTTTTATGATC 178
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
244 rTthYrsGlnPrOprhegluc1yAsnValrPheValrMetAspH1stYrH1sa 261
179 AAGAAGGTCGCCGT.....AATGATGAAGGTGAGCTCAAGTT 216
::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
261 spGlnlucysArGalaGlYrProG1u1ySpPheProAspSerAlaYgSer1le 277
217 GCCGGAATTCACCTTCATTCATTCATTCATTCATTCATTCATTCATTC 266
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
278 ...G1YleuThrValrPrOprheSerAlaYsAsnValrH1sArgYrAlrge 293
267 TCTGAATTCACGTCGATTTTGTATACACACACTGTTGCTTTCGTTTC 316
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
293 rleuAsnPrOlySglYllepHeValGlValSer1leValrPheMePheH 310
317 ATCGATTAATTTGTACCAAGTGTGATGATGATGATGATGATGATGATGAT 366
11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
310 isSer1leuPheMetHrYrThrAspGlnThrValYalYsAlGlnCysPhe 326
367 TACATGGAAGCTGATATAAACAGTTAGTCACACAGATTGAGGTATCTGAAT 416

```

```

seq_name: sp_invertebrate:Q19053

seq_documentation_block:
ID Q19053 PRELIMINARY; PRT; 484 AA.
AC Q19053:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)
DE E04D5.3 PROTEIN.
GN E04D5.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA MCMURRAY A.;
RL Submitted (OCT-1995) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,

```



```

58 erLeuser1leasnphelysthrngInglYAlaPhegluInlYhIsValTyr 74
160 GNGAAGGCTTTATGATCAGAGAGTGGCGTAATGATGAGGAGGAG 209
75 VALYSGlYhISYrSerMetLysHIScysArgThrAspAlaThrLeuL 91
210 TCAAGTGGCGGAATTTTCACCTTCATTTGATTCAGCAATGTTGGCGTA 259
91 usergInValAsnLeuThrValSerTyrSerIaLcysAspValIleArg 108
260 CACGATCTCTGAATCCACGCTGATTTTGTACAAACAACCTGTGCATT 309
108 InArgSerAsnProLysGlyIleMetMetThrAlaThrIleIleIle 124
310 TCGTTTCATCCATTTATTTGTACAAAGTGTACGCGATGCGATGAGTAC 359
125 SerPheHISProMetPheIleThrLysIleAspLysSerTyrLysVal 141
360 ATGCTTTTACATGGAAGCTGATTAACACAGTTAGTGCACAGATTGAGTAT 409
141 nCysPheTyrAlaGluAlaGlnLysThrValThrGlnGlnLeuAsnVal 158
410 CT..... 411
158 spIleAlaLysGluGlnGlnLysLysIlePheValMetValGlyAspGlu 174
412 .....GAATCACAACCTGCTTTCAACCTCAAT..... 441
175 GlUGlYglYThrValSerHISThrThrGlyAspGlnLysLysLeuHISL 191
442 .....GTCCGAGATC 451
191 sLeuAsnAspProSerThrGluGluValArgIleSerTyrAsnValProLeu 208
453 CAGTATGCCGTATGAAATTTGGATGGTGACACACCGGTCAACCACT 501
208 roAspCysLysTyrAlaValLeuThrGluSerLysThr...GluGluVal 223
502 CAATTGCTATCATTTGTCAGCCAGTATTATCAATGATGACATGCGATTC 551
224 AlaPheAlaThrValGlyGlnIleValIleValIleValIleValIle 240
552 T.....GAACCGTTGATCATTTTGGCGCGGTGTGCATTCCTGCT 592
240 aProGlyGlnAsnGlnThrSerProPheCysValThrValHISerCysA 257
593 TTGTGATGATGATGATACGCTGATCTGTAAGTAATCTAATGCTGATGA 642
257 snValLysAspGluThrGlyLysGluValGlnIlePheAspGluAsnGly 273
643 TGTGCTCTGTAATAATATTGCTAATAATATTGGAATATCCAAAGATTT 692
274 CysAlaIleValAspLysTyrIleuIleAsnAsnLeuGluTyrSerSerAsp 290
693 AATGCTGGCCAAAGACTCAGCTATACAAATATGCGGATCGA 735
290 unThrGlyGlnLeuSerGlnValCysSerTyrThrValArg 304

seq_name: sp_invertebrate:Q22164

```

```

RA LENNARD N.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., CORSEY T., COULSON A.,
RA CRAXTON M., DEAR S., DG Z., DURBIN R., FAVELLO A., FULTON L.,
RA JONES M., GREEN P., HARKINS T., HILLER L., JIER M., JOHNSTON L.,
RA JARDNER A., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFEEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
DR EMBL: Z66565; CA91480.1; -.
SO SEQUENCE 384 AA; 43964 MW; 9FEFF44 CRC32;

```

```

alignment_scores:
    Quality: 546.50      Length: 394
    Ratio: 2.240        Gaps: 9
    Percent Similarity: 61.929    Percent Identity: 31.218

```

alignment_block:

US-09-323-427-3 x Q22164 ..

Align seg 1/1 to: Q22164 from: 1 to: 384

```

13 CTTATGCTTCTGTACACTACACTATGATGCTATTCATTCGGT 62
10 LeuValGlnPheThrTyrPheValLeuGlyPheSerAlaIleGlnAs 26
63 TGACATGCTGTGCAAGGTGAGCCAGCAAAATGATGTGACCACTTCAA 112
26 pAspAsnGlnLeuIleGlyGlnProGluIleGlnCysAsnAlaAspThr 43
113 TACACATCAATTTTAATACAGTAATGATTCGAAAGAGACATTTATGTC 162
43 LeAspMetGlnPheArgThrArgLysGlnPheAsnGlyLysValTyrVal 59
163 AAGGCTTTATGATCAGACAGAGTGGCGT.....AATGA 197
60 LysGlySerTyrAsnArgProGluCysArgValAspTyrSerThrLysAs 76
198 TGAAGGTGAGCGTCAAGTTGCCGAATTCACCTTCATTTGATTCATGCA 247
76 pGlnPheGlyArgProValGlyGlyIleLysLeuAsnHISGlyAlaCysA 93
248 ATGTTGCCGCTACACAGATCTCGAATCCACGTGTATTTTGTACACACA 297
93 snMetAspArgGlnArgMetIleAlaProGluGlyMetMetPheSerThr 109
298 ACTGTTGCTATTCGTTTCATCCATATTGTTACCAAGTGGATGCGC 347
110 ValLeuIleIleSerPheHISProLeuPheLeuThrArgMetAspLysAl 126
348 ATATCGAGTACATGCTTTTACATGGAAGCTGATAAACAGTTAGTGAC 397
126 eTyrHISIleArgCysMetLysGlnLalaIalaArgThrValIleAla 143
398 ACATTTGAGTATCTGAATACACATGCTTTTCAAACTCAAAATTTGCCG 447
143 LaIleAspValSerAsnLeuProThrGluSerValGlnSerAspLeuPro 159
448 ATGCGAGTATGCCGTTATGAATTTTGGATGGTGACCAACCGGCAACC 497
160 MetProThrCysSerTyrThrIleArgArgAspGlnLeuAspGlyProI 176
498 ACTTCATTTGCTATCATTTGCTACAGCAAGTTATCATTAATGACATGCG 547

```

```

.....||| ::||:: || ||::||| |||
176 eLeuLysTyrAlaLysValGlyAspGlnValAlaHisArgTrpGlnCysA 193
548 ATTCTGAACCGTTGATTAATTCTTCGGCGGCTGGCATTCCTGCTTGTGC 597
|||||
193 sPerGlu.....AspTyrGlyLeuLeuValHisSerCysTyrVal 206
598 GATGATGCTGAACGGTGAATCTGTGAATTTCTAATGCTGATGATGCTGC 647
|||||
207 GluAspGlyGlnGlyLysGlnMetIleIleAspGlnArgGlyCysH 223
648 TCTGTGAATATTGCTAATATTGATATTCACAGCATTTA...A 694
|||||
223 sTrhAspArgLeuLeuLeuGlyAspProThrTyrValGluAlaLeuAsnM 240
695 TGCTGGCCAGAGAGCTCACTATACAAATATGCGATGCTACAGACTT 744
|||||
240 eLalArgTyrArgGlnSerPheValPheLysPheAlaAspArgIleAlaVal 256
745 TTCTATCAATGCCAGATCACTATTACCATTAAGAACCAATAGCAATG 794
|||||
257 ArgPheGlnCysGlnIleArgLeuCysLeuLysAsp...AspGlyGlyC 272
795 T.....GTTCGACACAGATGTTCAAGACCACAGAGATTGGAGCTG 835
|||||
272 sAspGlyIleThrProProMetCysSer.....P 282
836 TTTAAACAGGTGGTGGCGGAGCAAAACCTGCTGACGTGGCAACTTCT 885
|||||
282 heLysAspAlaAsnSerAsnGlnLysGlnIleValLysArgAsnAlaArg 298
886 TTACTCAAGAAAGATCTGCAGAACCGAGAAATATCATGATGATAGCAAC 935
|||||
299 ThrPheLysProLysAspAlaAsp.....MetPheSerGlnThrVal 312
936 TGTATATCAACACCCCTTGAATTTAGCATGATATCAAGCTTTGCCAGTTG 965
|||||
312 LTrpValMetAspLysGlnAsnGlyAspSerThrSerAlaGlnAlaLag 329
966 ATTATGACATCCGCTGCATCTTCTCAACATTAATGACACCTGTAATACTT 1035
|||||
329 LuIleArgGlu..... 332
1036 GCTGCAGTACAAAGATGGAATTCGATGTCACCA.....TTTGG 1073
|||||
333 ...LeuAspProGlnThrIleCysLeuAlaProLysLeuValValAla 348
1074 CTTCGCAATGTTATGGGTTTAAGCATTTGATGTCGCGCGCATTA 1123
|||||
348 eValLTrpPhePheThrPheIlePheValLeuPheValThrThrIleLeuV 365
1124 TTACATTTTCGTTAAATTTCTGTCACAAATCAG 1155
|||||
365 aLValValLTrpHisArgTyrCysLysAsnGlu 375
seq_name: sp_invertebrate:09XVN2
seq_documentation_block:
ID 09XVN2 PRELIMINARY; PRT; 610 AA.
AC 09XVN2:
DT 01-NOV-1999 (TREMBlrel. 12. Created)
DT 01-NOV-1999 (TREMBlrel. 12. Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12. Last annotation update)
DE F53B6.6 PROTEIN.
GN F53B6.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WHITE S.:
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

```

```

RT elegans."
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SANDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMER E., STADEN R., SULTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHDLMAN P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL Nature 368:32-38(1994).
DR EMBL; Z81086; CAB03116.1; -.
SQ SEQUENCE 610 AA; 68585 MW; D793181D CRC32;

```

```

alignment_scores:
  Quality: 511.00      Length: 270
  Ratio: 2.762         Gaps: 6
  Percent Similarity: 68.519   Percent Identity: 37.407

```

```

alignment_block:
US-09-323-427-3 x 09XVN2 ..

```

```

Align seg 1/1 to: 09XVN2 from: 1 to: 610

```

```

28 ATACACATATTGCAATGCTTATTGCG.....ATTCGGTGAACAATGG 71
||| ||||| ||||| ||| ..... |||||
61 ThrIleLeuIleLeuLeuSerThrPheSerCysPheGlnIleGlnAsnG 77
72 TGTGCAAGTGAAGCAGAAATTGAATGTGACCAACCTCAATATACATGA 121
||| ||||| ||||| ||| |||||
77 yValValGlyLysProGlnValAlaPheCysGlyIleAspTrpIleArgVal 94
122 ATTTTAATACAGTAATGCATTCGAGAGACATGTTATGTAAGAGCTTT 171
:: ||||| ::||| ::||| ::||| |||
94 yValAsnThrGlnHisProPheAsnGlyArgIleTyrValAspGlyGlu 110
172 TTATGATCAAGAGGTTGC.....CGTAATGATGAAGTGAAGTCATCACT 215
|||||
111 SerAspLysGlnHisCysValGlnHisSerAlaAspAlaHisSerSer 127
216 TGCCGGAATTTCACTTSCATTTGATTCATGCAATGTTGCGCGTACACAT 265
|||||
127 oGlnGlnPheThrIleProIleGlyAlaCysAsnMetArgArgGlnArg 144
266 CTTGCAATCCAGCGTATTTTGTGAACAACAAGTGTGATTCGATTCGTT 315
::||| ::||| ||||| ||| |||
144 TrIleuHisProArgGlyIleSerPheSerPheThrMetIleIleThrSer 160
316 CATCCATTAATTTGTTAACAAGTTGATCGTCATATGCAATCAATGCTT 365
|||||
161 HisProPhePheValThrGlyMetAspArgAlaPheSerIleArgCysP 177
366 TTACATGCAAGCTGATTAACAAGTTAGTGCACAGATTTGAGGATTCGAA 415
|||||
177 ePheLeuGlnIleLysGlyLeuAsnLagIleLysValGlyThrL 194
416 TCACAACCTCTTTCAACACTCAAAATTTGCCAGATGCGAGTACCGCTTA 465
|||||
194 eValArgProGlnHisValAspGlnGluTyrSerLeuProValLysAlaTyr 210
466 GAAATTTTGGATGTGAGCAACCGGTGAACCACTTCAATTTGCTATGAT 515
|||||
211 HisLeuLysAsp...GlyTleGlnGlyHisValLeuArgPheAlaGlnVa 226
516 TGTGACGCCAGTTTATCATTAATGACATGCGATTCGTAACACCGTTGATA 565

```

[illegible]

1105 TTGATTGCTGCCGTCATATTATTCATTTCG 1134
 ||||| :|||:|||||:
 352 Leuileuilethrllevalserleuthr 361

seq_name: sp_invertebrate:Q22680

seq_documentation_block: PRT; 609 AA.
 ID Q22680 PRELIMINARY;
 AC Q22680;
 DT 01-NOV-1996 (TREMblrel. 01. Created)
 DT 01-NOV-1996 (TREMblrel. 01. Last sequence update)
 DT 01-NOV-1999 (TREMblrel. 12. Last annotation update)
 DE T22C8.7 PROTEIN.
 GN T22C8.7.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilitida;
 NC Rhabdilitina; Rhabdilitoidea; Rhabdilitidae; Peloderinae; Caenorhabdilitis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA THOMAS K.;
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KESHAV J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULISTON J.,
 RA THEBERRY-MEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTECK L., WILKINSON-SPROAT J., WOHLDMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans".
 RL Nature 368:32-38(1994).
 DR EMBL; 249071; CAA88879.1; -.
 DR HSSP; P03000; 1TIF.
 SQ SEQUENCE 609 AA; 68882 MW; 0BA2EE0C CRC32;

alignment_scores:
 Quality: 453.50 Length: 287
 Ratio: 2.399 Gaps: 11
 Percent Similarity: 65.854 Percent Identity: 37.282

alignment_block:
 US-09-323-427-3 x Q22680 ..

Align seg 1/1 to: Q22680 from: 1 to: 609

3 GATGATTCGCTTATGCTTCTGCTACACTTATTCATTGCTTTATT 52
 |||:|||||:||||| ||| |||||:|||||: ||
 20 AspAsnSerAsnIyrAsnPhn...TyrTyrThrHis.....Me 31
 53 CGATTCCGGTTGACATGCTGTCGAGGTGACGAGAAATGTAATGGA 102
 :|||: ||| |||||:|||||:|||||
 31 tglu.Pro.....ProLysLeuGluCysGly 39
 103 CCAACTTCATATACATCAATTTTAATACAGTAATGATTCGAGAGACA 152
 :|||: |||:|||||:|||||:||||| |||||
 40 serGluGlyIleArgLeuHisIleAsnProThrGlyThrPheGlyGlnH 56
 153 TGTATTGATGGAAGTCTTATGATCAAGAGGTGCGGTAATGATGAAG 202
 |||||:|||||:|||||:|||||: |||: |||: |||:
 56 svAlIyrValArgIlyPhePheProGlnThrValCysHisLeuAsnIyrC 73
 203 GTGAGAGT.....CAAGTGGCCGAATTCACATTCATT...GAT 240
 |||: |||: |||: |||: |||: |||: |||: |||:
 73 ysthrArgLeuThrAsnArgProIleValMetAspLeuProPheArgGly 89
 241 TCATGCAATGTGTGGCGGTACAGATCTGAAATCAACGTGTAATTTTGT 290

||||| ||| |||:|||||: ||| :|||
 90 ProCysAsnValArgArgValArgAsnValAlaProProSerIleSerIy 106
 291 AACACAACTGTTGTCATTTCGTTTCATCCATTATTTGTTACCAAGTTG 340
 :|||:|||||:|||||: ||| |||||:|||||: ||
 106 rAspValThrValIleIleGlnHisIleProLeuPheValThrSerPhe 123
 341 ATGCTCATATGATGATACAAATGCTTTACATGGAACCTGATAAAGATT 390
 |||:|||||:|||||:|||||: |||: |||: |||: |||:
 123 sPlysAlaIyrArgLeuAsnCysIleIyrArgGlnGlnGlnSerThrLeu 139
 391 AGTCACAGATTGAGGTATCTGAAATC...ACAAGCTTTTCAAACTCA 437
 :|||: |||: |||: |||: |||: |||: |||: |||:
 140 GlnGlnArgIleAsnValSerAspIleProSerThrAlaLeuGlnSerIy 156
 438 AATTGTCGCCGATGCCAGTATGCCGTTATGAAATTTGGATGTGACCAA 487
 : |||: |||: |||: |||: |||: |||: |||: |||:
 156 sAsnAlaProLys.....CysArgTyrAspValLeuSerGlySerLeuA 171
 488 CCGGTCAACCACTTCAATTGCTATCATGTCGTCAGCCAGCTTATCATAA 537
 :|||: |||: |||: |||: |||: |||: |||: |||:
 171 sngIlyProValValArgPheAlaAsnValGlyAspValValValHisLys 187
 538 TGGACATGGGATTCGAAACCGTTGATACCTTCTGCCGCTTGTCATTTC 587
 |||||:|||||: ||| ||| |||||:|||||
 188 TrpThrCysAspSer.....AspArgPheGlyPheValValHisSe 201
 588 CTGCTTTGTCGATGATGTAACGCTGATACCTGTGGAATTTCAATGCTG 637
 ||| ||| ||| :|||: |||: |||: |||:
 201 rCysValValArgAspGlnSerIlyLysAspPheGlnPheIleAspGlnA 218
 638 ATGATGTCGCTCTGTAATAATATTGCTAAATATTGGAATTCACACA 687
 |||||:|||||: ||| |||: |||: |||: |||:
 218 rglIyLysValThrAspPheSerLeuPheProGlnValSerIyrSerAsp 234
 688 GATTTAATGCTGCGCAAGAACT...CACGATATACAAATATGCGGATGC 734
 |||||: |||: |||: |||: |||: |||: |||: |||:
 235 AspLeuLysSerAlaPheThrAlaValArgAlaPheArgTylAlaAspGl 251
 735 ATCACAGCTTTCTATCATATGCCAGATCAGTATTTACCATTAAGAACA 784
 : |||: |||: |||: |||: |||: |||: |||: |||:
 251 nValMetValHisPheSerCysGlnIleThrThrCysGlnLysGlnGlnA 268
 785 ATAGC...GAATGCTTTCGACACCATGTTACAGAACCAAGAGATTGCGA 831
 |||: |||: |||: |||: |||: |||: |||: |||:
 268 sngIyCysGlnGlyIleSerProIleCysArgProMetAspLeuGly 284
 832 GCTGTTAA 840
 :|||: |||: |||: |||: |||: |||: |||: |||:
 285 ProIleLys 287

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 16, 2000, 04:52:00 ; Search time 4841.33 Seconds

(Without alignments)
~728.156 Million cell updates/sec

Title: US-09-323-427-3

Perfect score: 1161
Sequence: 1 atgatgatcgtctattgc.....ttcgtccaatcagaagca 1161

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : GenEmbl.*

1: gb_bal:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vl:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vl:*
32: gb_htgl:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_bal:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*

45: gb_htg7:*
46: em_htgl:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	339.2	29.2	25823	34 CEF22B5	250044 Caenorhabdi
2	323.8	27.9	34831	34 CEC47G2	249125 Caenorhabdi
3	315	27.1	2310	34 CELCUT1	M55997 C.elegans c
4	261.4	22.5	724	35 AF125580	AF125580 Wuchereri
5	249	21.4	5792	34 ALU73005	U73005 Ascaris lum
6	217.2	18.7	2584	34 MAMCUP1	X96677 M.attellia
7	180.6	16.4	637	34 BMA012617	AJ012617 Brugia ma
8	156.6	13.5	358	34 BPA012618	AJ012618 Brugia pa
9	107.4	9.3	39478	34 CEF53F1	281088 Caenorhabdi
10	83.8	7.2	31536	34 CEF04D5	264496 Caenorhabdi
11	74.8	6.4	32412	34 CEF53B6	281086 Caenorhabdi
12	67	5.8	39736	34 CER07E3	249207 Caenorhabdi
13	64.2	5.5	110000	32 CER39A1_3	Continuation (4 of
14	62.6	5.4	36306	34 CEM142	273428 Caenorhabdi
15	62.2	5.4	36069	34 CEM265	281143 Caenorhabdi
16	56.2	4.8	25284	34 CEM01A8	271267 Caenorhabdi
17	54.8	4.7	28396	34 CELF10E7	U41264 Caenorhabdi
18	53.6	4.6	40265	34 CEM28	249911 Caenorhabdi
19	52.8	4.5	34368	34 CEM06D12	282073 Caenorhabdi
20	51.8	4.5	40600	34 CEF20D1	278542 Caenorhabdi
21	51.8	4.5	164548	32 CEF70D2	AL008880 Caenorhab
22	49	4.2	23953	35 CELK06A1	U23449 Caenorhabdi
23	49	4.2	31731	35 CELK03H9	U21318 Caenorhabdi
24	47	4.0	45713	35 CELB0511	AF067608 Caenorhab
25	46.8	4.0	27770	34 CEF22C8	249071 Caenorhabdi
26	43	3.7	164399	34 PFMA13P6	298551 Plasmodium
27	42.6	3.7	165302	41 AC009451	AC009451 Homo sapi
28	41.2	3.5	37872	34 CEF04F8	265665 Caenorhabdi
29	41.2	3.5	102195	32 CEF71H9	AL021575 Caenorhab
30	41.2	3.5	110000	32 CEF102E5_0	AL022276 Caenorhab
31	40.2	3.5	82652	8 ATAC007020	AC007020 Arabidops
32	40.2	3.5	95713	8 AF085279	AF085279 Arabidops
33	40	3.4	9901	2 U67530	U67530 Methanococc
34	39.8	3.4	12976	34 CEF53H1B	AL132851 Caenorhab
35	39.8	3.4	110000	32 CEF53H1_2	Continuation (3 of
36	39.4	3.4	5214	35 AF090533	AF090533 Dictyoste
37	39.4	3.4	6100	34 AB017910	AB017910 Dictyoste
38	39.2	3.4	91894	11 AC005739	AC005739 Homo sapi
39	39.2	3.4	124181	44 AF207067	AF207067 Homo sapi
40	38.8	3.3	5038	2 SGU40026	U40026 Streptococc
41	38.8	3.3	35955	34 CEF53B7	272510 Caenorhabdi
42	38.2	3.3	2296	35 AF044022	AF044022 Eurytides
43	38	3.3	19000	33 AC004479	AC004479 Homo sapi
44	37.8	3.3	1310	7 EGU21005	U21005 Euglena gen
45	37.8	3.3	11163	2 AE001500	AE001500 Helicobac

ALIGNMENTS

RESULT 1
CEP22B5/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS

CEP22B5 25823 bp DNA
Caenorhabditis elegans cosmid F22B5, complete sequence.
250044.1 GI:899234
HTG; Cuticulin; Elongation factor; GTP-binding ADP-ribosylation

INV 02-SEP-1999

factor: Phenylalanyl-tRNA synthetase: RNA binding.
SOURCE
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditiidae; Peloderiinae; Caenorhabditis.
REFERENCE
AUTHORS Wilson,R., Alnscough,R., Anderson,K., Baynes,C., Berks,M., Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J., Coulson,A., Craxton,M., Dear,S., Du,Z., Dublin,R., Favello,A., Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M., Johnston,L., Jones,M., Kershaw,J., Kirsten,T., Laister,N., Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B., O'Callaghan,M., Parsons,J., Percy,C., Rikken,L., Koopra,A., Saunders,D., Showkeen,R., Smaildon,N., Smith,A., Sonhammer,E., Stenden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M., Vaughan,K., Waterston,R., Watson,A., Weinstock,L., Wilkinson-Sproat,J. and Wohldman,P.
TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2 (bases 1 to 25823)
AUTHORS Sims,M.
JOURNAL
TITLE Direct Submission
Submitted (13-JUL-1995) Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rwenematode.wustl.edu
COMMENT Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.
For a graphical representation of this sequence and its analysis see:-
http://webace.sanger.ac.uk/cgi-bin/display?db=wormace&class=Sequence&object=F22B5
Current sequence finishing criteria for the C. elegans genome unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones.
EXCEPTIONS are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone F22B5. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
The true left end of clone F22B5 is at 24607 in sequence 266522.
The true right end of clone F22B5 is at 16002 in sequence 266523.
The true left end of clone F14E5 is at 4610 in this sequence. The start of this sequence (1..101) overlaps with the end of sequence 266522.
The end of this sequence (25723..25823) overlaps with the start of sequence 266523.
FEATURES
source
1..25823
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="II"
/clone="F22B5"
411..1471
/gene="F22B5.2"
join(411..462,584..663,745..1127,1216..1471)
/gene="F22B5.2"
/note="similar to RNA binding protein; cDNA EST yk39008.5 comes from this gene; cDNA EST yk499d6.3 comes from this gene"
/codon_start=1
/protein_id="CA90354.1"
/db_xref="GI:3876226"
/db_xref="SWISS-PROT:Q19706"

/translation="MAPAPEVSWAAVEEDNADPHIEGADGTRFETAFTVEVDGRMK
VTVQFKYINKRPVAVADRKMKVFGSGCGPAPQVATTVAAVEVMOFTFRNKGSD
ILDVQEDKQATKTSTRCHCKNDHSTICPKVAMQIDEDADADKTEKDMAG
MRPGQIDRNRSDENCRVNNLPDQNNEDLRLDFGIGVIRIFLAIADKVTGLPFG
FAFYTFPSRDDAAAIAMLNDIRYHNVKWKEMTRPSN"
1870..2668
/gene="F22B5.1"
join(1870..2158,2209..2387,2534..2668)
/gene="F22B5.1"
/note="similar to GTP-binding ADP-ribosylation factor"
/codon_start=1
/protein_id="CA90353.1"
/db_xref="GI:3876225"
/db_xref="SWISS-PROT:Q19705"
/translation="MGFLIKRKQARERERILILGLDNAGTKTLKKFLDEPTDTI
EPTGFDTKTVHFKDFQDLNMDVGSGSLKRWKYNFESTALIMVWSSPRELLCC
SEELKILGEIRLGA SILVYANKSDLPAGADVIAQVIOVSSVYSSEKKCFVLVDL
HSISHHMKTRFSCCALSGDRLVQAMTILCDVGSRLTLD"
complement(2833..4393)
/gene="F22B5.3"
complement(join(2833..3042,3088..3700,3823..4048,
4273..4393))
/gene="F22B5.3"
/note="similar to cuticulin; cDNA EST EMBL:T01970 comes
from this gene; cDNA EST CEESX90RB comes from this gene;
cDNA EST EMBL:T00412 comes from this gene; cDNA EST
yk386c7.3 comes from this gene; cDNA EST yk386c7.5 comes
from this gene; cDNA EST yk504b12.3 comes from this gene;
cDNA EST yk543g2.3 comes from this gene; cDNA EST
yk652b8.3 comes from this gene"
/codon_start=1
/protein_id="CA90355.1"
/db_xref="GI:3876227"
/db_xref="SPTREMBL:O19707"
/translation="MARYSGIGLGLLVASYSATPVNNVGEPEVECGPSTVNTN
TRNFBGHVYVKGLFDQECNDGCKQVAKIELPPTCNVARTRSINPQGVYTTTV
VVSFHPQVTKVDRAVYVQCEYMEADRTVSTQIEVSDLTTFQVAVPPICKEILL
GGPGEVQVQFATIQOQVYHKWTCDSERVDFTCAVHSGTVDGNGDTVQILDENGCAL
DKFLNNLVEPTDLMAGQEAHVYKYADRSQIFYCQISTYVKEPNEBACAPTCSEPG
FGAYKQNONAOFKREAPVMEINLQVRAELTFLVEGLENLPSSLQAOALVNS
REIGEDSPROELICSSPHISVYVFLDTITVYFIMVYSRMVSDMKQ5AC"
9691..10234
/gene="F22B5.4"
join(9691..9890,9940..10234)
/gene="F22B5.4"
/note="cDNA EST yk304c8.3 comes from this gene; cDNA EST
yk304c8.5 comes from this gene; cDNA EST yk594h3.3 comes
from this gene"
/codon_start=1
/protein_id="CA90356.1"
/db_xref="GI:3876228"
/db_xref="SPTREMBL:O19708"
/translation="MSLSKAPRKMTMFPVLSRNASTSNVSPARIOLKRAPEGHQY
SRNMSRDPREYKVAIQGDTFYQFLVRLGHAIEVYFLVLAWEFLFCSASVWSG
KAETLIDRSNSKAPWDERLRLDYMKPVAFLDGRTRRCELMEDLODEMLAARK
RGTR"
complement(10849..12154)
/gene="F22B5.5"
complement(join(10849..10984,11033..11343,11397..11521,
11578..11832,11917..12154))
/gene="F22B5.5"
/codon_start=1
/protein_id="CA90357.1"
/db_xref="GI:3876229"
/db_xref="SPTREMBL:O19709"
/translation="MNCQOSASKVRFLVCGFNKASRCSSEKLIPKDLIKYKNPS
EQIKQFKQLVGFNDIRLLETKMDNQYLILCHKAKMSIALACTYFNLSNVMSITY
SDSTVLSGTSITYKSLNRSIPDQTFGLSYQPDNTYNLPHLMHEADTFCKSVKNS
GCVDSKYAAVMKRVLHTKVRGPITYVYKVPDEVEPELEFDVYKFSVPDEHVYK
YNSIPEYERKEYLAKFKNGVGAAYNKGVEYIGLGVTESSGCEIGPTCDKNK
AAQAIFOSITODLPMKDLKDYLVRCSDKPEDSATWIRPFLRSHEMTPFAHIKFNVI
PDGLNFSKVFVSSNP5NAP"
complement(12240..14032)

TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*

JOURNAL Nature 368 (6466), 32-38 (1994)

MEDLINE 94150718

REFERENCE 2 (bases 1 to 34831)

AUTHORS Palmer, S.

JOURNAL Direct Submission

COMMENT Submitted (27-APR-1995) Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or r.w@nemastode.wustl.edu

CDS Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information.

For a graphical representation of this sequence and its analysis see: -

http://webace.sanger.ac.uk/cgi-bin/display?db=wormacc&class=Sequence&object=C4762

Current sequence finishing criteria for the *C. elegans* genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones.

Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

IMPORTANT: This sequence is not the entire insert of clone C4762. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone C4762 is at 1 in this sequence. The true right end of clone T24F1 is at 34731 in this sequence. The true right end of clone T05B9 is at 12750 in this sequence. The end of this sequence (1..101) overlaps with the end of sequence 249129.

The end of this sequence (34731..34831) overlaps with the start of sequence 249912.

FEATURES

source Location/Qualifiers

1..34831

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="II"

/clone="C4762"

2562..4878

/gene="cut-1"

join(2562..2679,2992..3217,3898..4636,4687..4878)

/note="cut-1"

/note="similar to cuticulin; cDNA EST EMBL:Cl0444 comes from this gene; cDNA EST EMBL:C12078 comes from this gene; cDNA EST yk262c12.3 comes from this gene; cDNA EST yk262c12.5 comes from this gene; cDNA EST yk265h3.5 comes from this gene; cDNA EST yk278g8.3 comes from this gene; cDNA EST yk278g8.5 comes from this gene; cDNA EST yk439e8.3 comes from this gene; cDNA EST yk439e8.5 comes from this gene; cDNA EST yk380a11.3 comes from this gene; cDNA EST yk583c11.3 comes from this gene; cDNA EST yk380a11.5 comes from this gene"

/codon_start=1

/protein_id="CA88934.1"

/db_xref="GI:3875030"

/db_xref="SWISS-PROT:Q03755"

/translation="MTWPKITCLALVLSASAIPIVDNNVEGEPEEGCPNISTYVNFNTINPEGHVYVGLIDAGCRSDGEGROYAGLEPDEDSCTARTSLNKGVEFTYVIRSHPORYTKYDRAVYRIQCETMESDKYTSQIEVSDLTARQOVVPMKVCYKILLG GPGSQPIQFATIGQOVYHKWTCDSSETDTDFCAVYHSCIVDNGDVIOLNEECALDKELLNLEYPDIDMAGQEAHYKYADRQQLPYQOQIISTTKDPGSECARPTCSPPQGRKRRSGENGILDIVREINTLIDIMEGASPSAEEAALVSEESVRRRATSTGICLIP IGKSPFLIGITIVATLSATTFYVARPTSHKH"

15433..17108

/gene="C4762.2"

join(15433..15770,16205..16531,16772..17108)

/gene="C4762.2"

/note="similar to fork head domain protein; cDNA EST EMBL:T01618 comes from this gene; cDNA EST EMBL:D70725 comes from this gene; cDNA EST EMBL:D66805 comes from this gene; cDNA EST yk268f8.3 comes from this gene; cDNA EST yk268f8.5 comes from this gene; cDNA EST yk474g11.5 comes from this gene; cDNA EST yk637f8.3 comes from this gene; cDNA EST yk670a12.3 comes from this gene"

/codon_start=1

/protein_id="CA88935.1"

/db_xref="GI:3875031"

/db_xref="SPTR EMBL:Q18694"

/translation="MLTSMESILSTYRPLEPPKLEPREVTINEQVYDPRSNTRLSEPTSAVSLEHDLKFGESRRKRSLSGDEPTEDSDVPVAKARKNRHSTSSADSSDDAKDDDDDDSTSRKSMGSRKSSSHAKRPSTYLLAMSLNPERKLLSECEIINKFEYRKERPAMONSIRNHLSLNDFEVKARPGNPGNGYALRPNLDEGSSFLRRKRYKKNSDPMYHNSHHPRRPRLPGCGPRPRRPMANIPMLGHPNPAVNMARFIPQNDISOKLISMASTRIMPDAVPVSSGQKRISSSSPENSSSAVSKLSA"

complement(24417..24975,25043..25176,25237..25522,25668..25907,25967..26054,26106..26227,26277..26608,26681..26785,26846..26914)

/gene="C4762.4"

/note="similarly to Trichostrongylus colubriformis 11 kd secretory protein (Swiss Prot accession number P21937); cDNA EST EMBL:D3349 comes from this gene; cDNA EST EMBL:D37644 comes from this gene; cDNA EST EMBL:D36149 comes from this gene; cDNA EST EMBL:C11456 comes from this gene; cDNA EST EMBL:C13631 comes from this gene; cDNA EST yk338b10.3 comes from this gene; cDNA EST yk355b10.5 comes from this gene; cDNA EST yk301f4.3 comes from this gene; cDNA EST yk301f4.5 comes from this gene; cDNA EST yk295f2.3 comes from this gene; cDNA EST yk295f2.5 comes from this gene; cDNA EST yk550a1.3 comes from this gene"

/codon_start=1

/protein_id="CA88936.1"

/db_xref="GI:3875032"

/db_xref="SPTR EMBL:Q18695"

/translation="MGTISLAVQLEFVILFSLYLANKYSTRKNPIVTISTFGWYFLIVPVLPIQVATTFPHKGCNDORILNTTSPVAPVPCELPGGVDPDVFNLN RVYWSQQLTWLILPLOSIVYRAGNTITLCKIRAAVYINNALYAIYSLCFALITAMKGVASINTELVKIVVASNTWGLFVLVLLGLGALDEPSLHNGNRRLRKLKTFDIERLASESEAEENKDIYKRVVGLNDESGORAKRVTLISKFSVDVNDLNP SROYIDNANLDESGSCSEAKLISLHKRTIYAVQVLLNNAQMLVLRALFLEMLAS ESNQYNLELSNTECPVIGVRFREYTRQTPCRLIGIVTEMFEVFLSECTFVAVSY TLSPAATVEYASRFRHYKYQFVAFQIVYLITASVYTFERLOIYVHLDHPGVD ENSLIESAILLCRTPLICLNFPLSMIHMSITSNAKSFQIOTKMLGHIDVPIILA KGINVLPICILICLAIHYRVGAVYVHLNIGEDQFVEADENTNIMSGSLVIERN SIKRNSNRSORTQMTWNTSPGSSNAGNSTTSKFRSKNDEERPLDDEDEVEEST RISMSPTHERPSSGCFDDM"

30076..34049

/gene="C4762.5"

join(30076..30134,30184..30492,31050..31274,31321..31480,31534..31653,31701..31835,31889..32104,32160..32570,32639..32935,32998..33263,33263..33651,33802..34049)

/gene="C4762.5"

/note="cDNA EST CEMSB75F comes from this gene; cDNA EST EMBL:C13621 comes from this gene; cDNA EST yk227d8.5 comes from this gene; cDNA EST yk257e7.5 comes from this gene; cDNA EST yk319c4.5 comes from this gene; cDNA EST yk31d4.5 comes from this gene; cDNA EST yk335d12.5 comes from this gene; cDNA EST yk486d7.5 comes from this gene"

/codon_start=1

/protein_id="CA88937.1"

/db_xref="GI:3875033"

/db_xref="SPTR EMBL:Q18696"

/translation="MMAEGRRNNRKKSRDRDREGTSASDAYSSDSADDEDITQDDDDPMLNLNDSLDHNVEARDEPDELIVTIVSPNMPEDNEDEHEMKFEEDSGSR IATSSASEAVTQDGRSDSDSHTARSLIYTSRPNMSPAKLSKMTFSIPOMFAKGS ENSLMLRLKTDNFTLEBDVLYNERYVQSGRKGAELIYVITSRMKALLLSLNPKN TIDPFRKQOYRSPTASIELITRTGVDYQKQIVTTEETRKCLVDFLNKTPLNHVAQV FAKIMEDCLSRHEDLPESLLOETKFFDKCLRNINLGAIECLLEMLVRIPTTSGRTIV

Db	301	TTGCAGCTGCATTCACGACACACTGCTGCATTTAGAGTTTGAAGAAAAGAGACATGTGAAG	360
QY	911	cgaggaatacatcatgatgatgcgaactgacatatacaacaccttgaaattagcgatgataac	970
Db	361	ATGAAATATATTGTGATGATACGTACCGATCTTAATGCAATTGGATATAAATGAAGGTTTA	420
QY	971	aagcttgcgcagcttgatattaagtaacgctgactctg-----caacataatgacac	1024
Db	421	CCTCATTTACCAAAATGCCGTACGTATCATCATATCATTTGTTAGACATGAGATGATGATC	480
QY	1025	ctgcaataactctgcgcagtaacaaatggaaatctcatgatacacaattggcttcacatgt	1084
Db	481	CGGTATATTGTTGCACACATATGACACAGGTATATGATATATGATATACCGGTTTACACTTG	540
QY	1085	ttaatgggttaagaatcatgatgatgatgtgcgcgcgtcatatt	1125
Db	541	CCGGATGCTAATATTGTGTATGTGTCTGTTCGTACAAATT	581
RESULT	5		
ALU73005	5792 bp	DNA	INV
LOCUS	Ascaris lumbricoidea CUT-1-like cuticlin protein precursor		14-AUG-1997
DEFINITION	(ascut-1) gene, complete cds.		
ACCESSION	U73005		
VERSION	U73005.1	GI:1657624	
KEYWORDS			
SOURCE	common roundworm.		
ORGANISM	Ascaris lumbricoidea		
REFERENCE	Ascaris lumbricoidea: Eukaryotes; Metazoa; Nematoda; Secernentea; Rhabdittia; Ascaridida; Ascaridoides; Ascarididae; Ascaris.		
AUTHORS	1 (bases 1 to 5792)		
TITLE	Timinouni,M. and Bazzicalupo,P.		
JOURNAL	cut-1-like genes of Ascaris lumbricoidea		
MEDLINE	Gene 193 (1), 81-87 (1997)		
REFERENCE	97390131		
AUTHORS	2 (bases 1 to 5792)		
TITLE	Timinouni,M. and Bazzicalupo,P.		
JOURNAL	Submitted		
FEATURES	Submitted (01-OCT-1996) GA3, I.I.G.B., via G. Marconi, 10, Napoli, NA 80125, Italy		
Source	Location/Qualifiers		
	1..5792		
	/organism="Ascaris lumbricoidea"		
	/variety="sum"		
	/db_xref="taxon:6252"		
	766..881		
	/gene="ascut-1"		
	join(766..954,2851..3239,3583..4083,4499..5081)		
	/gene="ascut-1"		
	/product="CUT-1-like cuticlin protein ASCUT-1"		
	766..5081		
	/gene="ascut-1"		
	766		
	/gene="ascut-1"		
	/note="site of transsplicing by the SL1"		
	882..929		
	/gene="ascut-1"		
	join(882..954,2851..3239,3583..4083,4499..4693)		
	/gene="ascut-1"		
	/note="cuticlin gene; ASCUT-1: homologous to cuticlin proteins of other nematodes"		
	/codon_start=1		
	/product="CUT-1-like cuticlin protein precursor"		
	/protein_id="AA06664.1"		
	/db_xref="GI:1657625"		
	/translation="MCRVAVFLAFGLAIPVNDGVEPEIECGPSTIVNTENTN		
	PFGHAAVKGLDIOEGCRSDGGRGVAGISLPDSCNVAITRLNPGICFVTTVVS		
	FHPLEITKVRAYRVOCFVMEADKYVSTOIESEITTAFOGLIIVMPVCHVEILDGP		
	TGQPIQATIGOOVYHKMTCDSEVDFEFAVHVSCHPVDGSGDTIILNEGALDKY		
	LVNNLEPTDLMLAGEAHYVKIADRSOLFQOCITTIKFPNSCEPPTSEFGKRA		
	VRRGSLAPKRRRCQRLRIKKSGGDVNTLDVPTDFSAIDISDRALPMDLRHRR		

Query Match	21.4%	Score 249	DB 34	Length 5792
Best Local Similarity	69.7%	Pred. No. 7.7e-49		
Matches 352	Conservative 0	Mismatches 150	Indels 3	Gaps 1
mat_peptide	HARCOOVLSPANGICMSPGSEIFEMGLAVLAALAAVYVVSFKLRPOOKA"	join(930. .954,2851. .3239,3583. .4083,4499. .4690)		
	/gene="ascut-1"	/product="cut-1-like cuticlin protein"		
3'UTR	4694. .5081	/gene="ascut-1"		
polyA-signal	5064. .5070	/gene="ascut-1"		
BASE COUNT	1689 a	1254 c	1213 g	1636 t
ORIGIN				
Query Match	21.4% <td>Score 249 <td>DB 34 <td>Length 5792 </td></td></td>	Score 249 <td>DB 34 <td>Length 5792 </td></td>	DB 34 <td>Length 5792 </td>	Length 5792
Best Local Similarity	69.7% <td>Pred. No. 7.7e-49 <td></td> <td></td> </td>	Pred. No. 7.7e-49 <td></td> <td></td>		
Matches 352	Conservative 0	Mismatches 150	Indels 3	Gaps 1
463	ttatgaatttttgatgtytgagaccacccggttaacacgaatttcataatttgta 521			
Db	3576	ttttacagatttcgcagcgaggtcccaacggagacgcgatacgttcgcaacatcgctga 3635		
Oy	522	ggcagttatcatatgaatgacatgcgattctgaacccgtttgacttttgcgcggtt 581		
Db	3636	gcaagtatrrcccaagtgacatttgcgatttcggaacctgttgatcgttttgcgtcgt 3695		
Oy	562	ccattcccttgcttgcagatgcatgtaacggttgatagtctgtgaattctaatgctatg 641		
Db	3696	ttcactcattcgttgcgtttgatgatggcagcggtgatcagatccagatccacagagag 3755		
Oy	642	atgtgccttgtaataatttcttaataatttgtaatatccaacagatttaatgctg 701		
Db	3756	ttgtgcgcttgcaaaatattctctgamaaacctcgagttcccccacgatttgatggccg 3815		
Oy	702	ccaagaagctcagatatacaaatatgcgatacgcatacagcttttcataatgccaag 761		
Db	3816	ccaaagaagcgacgcttacaataatgcgacccgatacagcttcttatacagtgccagat 3875		
Oy	762	cagattaccatlaaagaacccaataagcgaatgctgcagacaatgctcagaaccaca 821		
Db	3876	tactattacacattaaaaagacccaaacagatggcccaagaccacattggcagtgaccga 3935		
Oy	822	aggtatcgagagctgttaaacaggtggtgc---ggaagcaaaacctgcgcagctcgca 878		
Db	3936	aggaattttggggctgtgcgtccaggtggctccatagggccgcaaaaacacggcgctgcca 3995		
Oy	879	acttcgtttacccaagaagaagatctgcagacccggaatatcatgttgatgtaacga 938		
Db	3996	actccgccttgatcaaaaagacgtgggggtgactatgacacacaccttgatgtagcacga 4055		
Oy	939	tatcaacacccttgaataatgcatg 963		
Db	4056	ctttcagcgccctcgacatttagctgat 4080		
RESULT 6				
NAMECUT1	2584 bp	DNA	INV	25-NOV-1997
LOCUS				
DEFINITION	M.artiella Mcut-1 gene.			
ACCESSION	X96677			
VERSION	X96677.1 GI:2648040			
KEYWORDS	cuticle protein; cuticlin 1; Mcut-1 gene.			
SOURCE	Meloidogyne artiella.			
ORGANISM	Meloidogyne artiella			
	Eukaryota; Metazoa; Nematoda; Secernentia; Diplogasteria;			
	Tylenchida; Tylenchina; Tylenchoidea; Heteroderidae;			
	Meloidogyninae; Meloidogyne.			
REFERENCE	1 (bases 1 to 2584)			
AUTHORS	De Giorgi,C.			
TITLE	Direct Submission			
JOURNAL	Submitted (15-MAR-1996) C. De Giorgi, Dipartimento di Biochimica e			
REFERENCE	2 (bases 1 to 2584)			
AUTHORS	De Giorgi,C., De Luca,F., Di Vito,M. and Lambertini,F.			
TITLE	Modulation of expression at the level of splicing of cut-1 RNA in			

	the infective second-stage juvenile of the plant parasitic nematode Meloidogyne artiellia							
JOURNAL	Mol. Gen. Genet.	253 (5),	589-598	(1997)				
MEMLINE	97218031							
FEATURES								
SOURCE	Location/Qualifiers							
	1..2584							
	/organism="Meloidogyne artiellia"							
	/db_xref="taxon:42426"							
	/clone_1lb="lamda7-2(1)"							
	/clone="pluc(3000) "							
	/clone="mp19(1500) "							
	join(<467..943,1139..1735,1825.>2052)							
mRNA	gene="Mtcut-1"							
	<467..943							
exon	/gene="Mtcut-1"							
	/number=1							
CDS	join(467..943,1139..1735,1825..2052)							
	/gene="Mtcut-1"							
	/codon.start=1							
	/product="cuticlin-1"							
	/protein_id="CA65452.1"							
	/db_xref="GI:2648041"							
	/db_xref="SPTRMBL:O18479"							
	/translation="MRKLFLAIGVFAINLAFTVALIPVDNGVEGEDEIECGPTSTIVNENRNPEGHVVYKGLFDGAGCSDEHGRQVACIELPPOSCNVARPDAPDKGVSTTVTISFHQPFTKYDRAYRNOCFEMEDKITYSADLESEITTOOTOVVPMPCYEITLEGALGPDIOPFATIAOOVYTSGETSETIDTFCAVNHSCVWDNGSDVOILNEGSCALDFLLNNLEYPPDLTAGOEAHVKYADRSQLPYCOQISTIKERPSGARPKCARPSGFVAHVAGAGACAGAPPMAAAATPPPAAAPAPAIAPARVQLRTORLLRKRRAPRGEOVLDTREFNALDISDKPFNLPAIRHHHLMENTLSAEPRRFLYGGQQVCMSSFRANHLIRVHGCHNGCSHMWVHSYLCFIYVFESNEIK"							
intron	944..1138							
	/gene="Mtcut-1"							
	/number=1							
exon	1139..1735							
	/gene="Mtcut-1"							
	/number=2							
intron	1736..1824							
	/gene="Mtcut-1"							
	/number=2							
exon	1825..>2052							
	/gene="Mtcut-1"							
	/number=3							
POLYA_signal	2518..2523							
BASE COUNT	761 a 639 c 559 g 625 t							
ORIGIN								
Query Match	18.7%	Score	217.2;	DB	34;	Length	2584;	
Best Local Similarity	56.8%;	Pred.	No. 2.4e-41;					
Matches	581; Conservative	0;	Mismatches	243;	Indels	198;	Gaps	2;
Oy	51	tccgattcccggttgacgaatggtgcgaaggtttagcagaagaattgaatgtggacaactc	110					
Dd	529	tgccattttccacgtgacacaaaggctgtggaagcagaccgaagattgatatgtggcccacactc	588					
Oy	111	aataacaatcaatttaataacagtaatgatcatctgaaagacatglttatgtgnaaagtct	170					
Dd	589	aatcacccgtcaattttcaacacacacgaacacctttcgagggcagcgtatgtgaaaggcct	648					
Oy	171	ttaigtatcaagaagtttgcgcftaatgataagaaagtgtgaacgtctaagtctgcgaatttcac	230					
Dd	649	citttgaccaagccggatgccccgttgagatgacacgcccgaacaagtgccgggaattgagact	708					
Oy	231	tccatttgtatcatcatcaatgtgtgcgctgatacaagatctggaatccaagtyglatlittgt	290					
Dd	709	tccgttgacacttttcgcatattgttccccgcacatgatct---gaaccacaagcagtgttct	765					
Oy	291	aacacaacatgttgatcatctcglttacatccattatgttaccaaaagtgtatgycatala	350					
Dd	766	ctccacaaagggcgacatctcttccatctcacaagtttgacaaagttgacccgggacct	825					

OY	351	tcgagtcacaaagcctttacatlggaagcgtgataaaagctagtgacacagatctgagtcac	410
Db	826	CCGCTGCCAGCTGCTTCAATATGGAAAGACAGACAGACGCTCTGGACACAGCTCGAGGTCTC	885
OY	411	tgaatccacaacatctcttccaacatcaaaatltgcgcgaatgcagatgcgctlatg	467
Db	886	CGAGATCACACACAGATTCACAGACCCAGGTGGTCCCTATGCTGTGTGCATAATATGAGGT	945
OY	467	-----	467
Db	946	AAGTAAAGAAGAAGACGACCTCTTTCAGTAATATCATATSCAAATCTTGAAATTTGGCAT	1005
OY	467	-----	467
Db	1006	AAATTCGAGTGTCTTAATTTCACATTTTGGTCTCTCTTACCACAGCTTTAAGATGCTCAA	1065
OY	467	-----	467
Db	1066	AATGTTTGTTTTTTAAACCTTTATATAGTAAAAAACAAACAAATATTTTTAAATGAATT	1125
OY	467	-----aaatttggatggtggaacaacggttcaacacagttcaatttgcatacat	515
Db	1126	TATTTCTTCCAAAGATCCTCGAAGGCGCCGCACTTGGGCAACCAATTCATTTGCGACCAT	1185
OY	516	tgctcagccagtttatcatataaaltggaacatgcagatltctgaaccggttgaatacttctgcgc	575
Db	1186	TGCCACAACAACTACACAAAGTGGAGCTGTGACTGTGAGACATGTGACATTTCTGTGGCC	1245
OY	576	ggtgttcattccctgcgtttgtgcgaatggttaagtgtaacgtgtatctgtggaatcttcaatg	635
Db	1246	GTTGTGTCAACAGCTGTTGTGTGATATGTCATATGGGACACCGTGCAAAATTTTGAACGA	1305
OY	636	tgaatgattgtctctcttbatataatcttgcataaataattgggataaccaagaataat	695
Db	1306	AGGAAGCTGTGCCCTGAGCAAGTTCTTGCTCAACAACACTTGGAAATACCAACGATTTGAC	1365
OY	696	ggtctgcccagaagaagctcaagtlacatacaataatgcgaatgcgaatcaacagctttctcatatg	755
Db	1366	TGCCGGCCAAAGAGCTCTACCTCTACAAATACCGGGATCCCTCTCAACTCTTCTACCAATG	1425
OY	756	ccagatcagatcttaccatataaagaaccaaatagcgaatgtgttcgacacaaatcttcga	815
Db	1426	CCAATCTAGTATTAACCTTCAAGSAGCCACACAGCTAGATGTGCTGCTCCAAAGTGTGCTGA	1485
OY	816	accacaagagatcggagctgttaaaaacaggtggtgcgcgagcaaaaacatctgtcagctgc	875
Db	1486	GCCAAAGCGGCTTCAACGCCCTCAAAAGTGGGCGCCGGTGGAGCTGTGCGGTGCCCCAC	1545
OY	876	gc 877	
Db	1546	GC 1547	
RESULT 7			
BMA012617	LOCUS		
DEFINITION	BMA012617	637 bp DNA	INV
ACCESSION	Brugia malayi cut-1 gene, partial.		29-JUN-1999
VERSION	AJ012617.1	GI:3858954	
KEYWORDS	cut-1 gene; cuticlin.		
SOURCE	Brugia malayi.		
ORGANISM	Brugia malayi.		
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;		
AUTHORS	Filiarioidea; Onchocercidae; Brugia.		
	1 (bases 1 to 637)		
	Lewis, E., Hunter, S.J., Telley, L., Nunes, C.P., Bazzicalupo, P. and		
	Devaney, E.		
TITLE	cut-1-like genes are present in the filarial nematodes, Brugia		
JOURNAL	panhagi and Brugia malayi, and, as in other nematodes, code for		
	components of the cuticle		
	Mol. Biochem. Parasitol. 101 (1-2), 173-183 (1999)		
MDLINE	99339397		

FEATURES	source
REFERENCE	2 (bases 1 to 637)
AUTHORS	Devaney, E.
TITLE	Direct Submission
JOURNAL	Submitted (06-NOV-1998) Devaney E., Veterinary Parasitology, University of Glasgow, Bearsden Road, Glasgow, G61 1QH, UK
source	Location/Qualifiers 1..637
gene	/organism="Brugia malayi" /db_xref="taxon:6279" /clone="cut-1"
exon	1..484 /gene="cut-1" 1..164 /gene="cut-1"
CDs	/number=1 join(<2..164,266..>484) /gene="cut-1" join(<2..164,266..>484) /gene="cut-1" /codon_start=1 /product="cuticlein" /protein_id="CA10074.1" /db_xref="GI:3858955" /db_xref="SPTREMBL:O96775" /translation="V1SFHPLSVTKVDRAYVOCGYEADKTVSTQIEVSEITTAFODT QIVMPVCRIYELDGGPTGDPIDGAPVYHKMTDSEIVDFCAVHSCFVDGDN GDPEILSADSCALDKYLLNLEXP"
intron	165..265 /gene="cut-1" /number=1 266..484 /gene="cut-1" /number=2
BASE COUNT	174 a 136 c 134 g 193 t
ORIGIN	
Query Match	16.4%; Score 190.6; DB 34; Length 637;
Best Local Similarity	68.3%; Pred. No. 4.4e-35;
Matches 334; Conservative 0; Mismatches 54; Indels 101; Gaps 1;	
Db 303	tgctcattcgcttaccatcctatctgtgtaccacaaattgatcgtgcatatcgaatcaatg 362
1	tgtaatttctctccatccactctgcgtcacaaaattgatgcagcatccaggtcactg 60
363	ctttacatlgagaagctlgataaaacagtttagtgcacagattgaaglatctgaaatcacaac 422
61	cttctatatgagaagctgacaaacagtcagaccacagattgagtgctggaatatcacaac 120
422	tgcctttcaaatcaaatctgtcccgatgcccagatagccggtatgaa----- 469
121	tgcttttcaaacctcaaatgttcccatccgtctgtgtgcatacgaagcatgtcttttga 180
469	----- 469
181	attttttgatttgattttatttatcaattcaatcaaatcaaatcaaatatttatca 240
469	-----atttgatgltggaaccaaccggtcaaccagtt 501
241	catagtacatgtgtgacaaatatattttagattttgagttgtgaccaaccgacagcaatt 300
502	caatttgcatacatgtgacgacgagttatataaaatgagatgcatgtcgaacggt 561
301	caatttgcaccattggccaaacacagtttatcaacaatgacacctgtgattccgaacggt 360
562	gatacttctgcgcggtgtgcatcctccgtcttgcgtatgcatgltgaaagcgtgatactg 621
361	gataccttctgcgcgagttgtccacacgctgttgatgatatggcaacggatgacggtg 420
622	gaaatctcaaatgctgatgatgctcttgataaaatctgtctaataatttggaaat 681
421	gaatttctgagtgacagatggttgctcttcataataatctgtcgaataatttggaaat 480

QY	682	ccacacgat	690	
			1	
Db	481	CCGACAGGT	489	
RESULT	8			
LOCUS	BPA012618	358 bp	DNA	INV
DEFINITION	Brugia pahangi cut-1 gene, partial.			29-JUN-1999
ACCESSION	AJ012618			
VERSION	AJ012618.1	GI:3858956		
KEYWORDS	cut-1 gene; cuticlin.			
SOURCE	Brugia pahangi.			
ORGANISM	Brugia pahangi			
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spirurida; Filarioidea; Onchocercidae; Brugia.			
AUTHORS	Lewis, E. Hunter, S.J., Telley, L., Nunes, C.P., Bazzicalupo, P. and Devaney, E.			
TITLE	cut-1-like genes are present in the filarial nematodes, Brugia pahangi and Brugia malayi, and, as in other nematodes, code for components of the cuticle			
JOURNAL	Mol. Biochem. Parasitol. 101 (1-2), 173-183 (1999)			
MEDLINE	99393937			
REFERENCE	2 (bases 1 to 358)			
AUTHORS	Devaney, E.			
TITLE	Direct Submission			
JOURNAL	Submitted (06-NOV-1998) Devaney E., Veterinary Parasitology, University of Glasgow, Bearsden Road, Glasgow, G61 1QH, UK			
FEATURES				
Source	Location/Qualifiers			
	1..358			
	/organism="Brugia pahangi"			
	/db_xref="taxon:6280"			
mrna	join(<1..191,350..>356)			
	/gene="cut-1"			
exon	1..191			
	/gene="cut-1"			
	/number=1			
gene	1..356			
	/gene="cut-1"			
	join(<1..191,350..>356)			
CDS	/gene="cut-1"			
	/note="unnamed protein product"			
	/codon_start=1			
	/protein_id="CAAI0075.1"			
	/db_xref="GI:3970656"			
	/db_xref="SPTREMBL:O96776"			
	/translation="PIPTATIGQGVYHKWICDSEYDTFCAVYHSCFYDGGNGDTVEI			
	LNADGADLDKLLNNLEPYDL"			
	192..349			
intron	/gene="cut-1"			
	/number=1			
exon	350..356			
	/gene="cut-1"			
	/number=2			
BASE COUNT	110 a 64 c 64 g 120 t			
ORIGIN				
Query Match	13.5%	Score 156.6	DB 34	Length 358
Best Local Similarity	87.7%	Pred. No. 4.6e-27		
Matches 171: Conservative	0: Mismatches 24: Indels	0: Gaps	0:	
QY	496	ccagttcaattgcatcatatgctgcagccagcttcatcataatgacatgcgcatctgaa	555	
Db	1	CCAATTCAATTGCTACCACTTGGCCACACAGTTTATTCACAAATGACCTGTGATTCGAA	60	
QY	556	accgttgatacttctgcgcggtgtgtccatctcctgttctgtcagatgatacgtgatat	615	
Db	61	ACCGTTGATACCTTTGGCAGAGTGTCCACATCTCGCTTTGTGATGATGGCAACGGTAT	120	
QY	616	actgttggaattctaaatgctgatagtgctctcttgataaataattgtctaaataattt	675	

Db	121	ACGGTGGCAATTTCTGAATGCAAGATGTTGTGCTCTCCGACCAATFACTCTGCTGAACAATTTG	180
Ox	676	gaataccaacagat	690
Db	181	GAATATCCACAGCT	195

RESULT	9
CEFS3F1/c	CEFS3F1 39478 bp DNA INV 02-SEP-1999
LOCUS	Caenorhabditis elegans cosmid F53F1, complete sequence.
DEFINITION	Caenorhabditis elegans
ACCESSION	281088
VERSION	281088.1 GI:1627965
KEYWORDS	HTG.
SOURCE	Caenorhabditis elegans.
ORGANISM	Caenorhabditis elegans
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
AUTHORS	1 (bases 1 to 39478) Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kersey, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roop, A., Saunders, D., Shownkeen, R., Skilton, R., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterson, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J., and Wohlman, P.
TITLE	2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
JOURNAL	Nature 368 (6466), 32-38 (1994)
MEDLINE	94150718
REFERENCE	2 (bases 1 to 39478)
AUTHORS	Burton, J.
TITLE	Direct Submission
JOURNAL	Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rwenematode.wustl.edu
COMMENT	Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information. For a graphical representation of this sequence and its analysis see: - http://webpac.sanger.ac.uk/cgi-bin/display?db=wormacc&class=Sequence&object=F53F1 Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note. IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. This sequence is the entire insert of clone F53F1. The true right end of clone M04612 is at 21759 in this sequence. The start of this sequence (1..101) overlaps with the end of sequence Z81103. The end of this sequence (39379..39478) overlaps with the start of sequence AL021448. Location/Qualifiers 1..39478 /organism="Caenorhabditis elegans" /db_xref="taxon:6239" /chromosome="v" /clone="F53F1" complement(14923..16486) /gene="F53F1.1" complement(join(14923..15347,15663..15774,15835..15984,16033..16271,16318..16486)) /gene="F53F1.1" /note="Predicted using Genefinder; similar to cuticlin" /codon_start=1

gene
protein_id="CAB03124.1"
/db_xref="GI:3877456"
/translation="MDNKEIRRTLEDYEPKGLIESSKYDCHELDMLVLQDLIATL
AVNPPIQNSLGVDOVECDSPRTISVQIKTEPEFGVILFVKPFAEEVCTSGTSL
AFLEIEIGCGALROVRVNLNPKGLAVRTITISFHFYITKRVNLAIKLVESQV
ANIIISVDETSITSYNNVNLMPCTQOILISGCPGPEVRLIGQOVYHOMKCDNGM
ALLIEDSPSCVYVHTCSYVDGSGTSEFLDNGSGSIDKFLISNLEYPGLADGSAHV
YKFAORDALFFPOCQISITVKEPDDCEVPRICDVGCGAPVYGPPTGKISNLYNI
QKNSLTFFPKSPTMYVVGITTYHNKAIOKL"
17396..18524
/gene="F53F1.2"
CDs
join(17396..17599,17683..17926,17977..18146,18197..18286
18339..18524)
/gene="F53F1.2"
/note="similar to aldehyde reductase; cDNA EST yk473d8.3
comes from this gene, cDNA EST yk473d8.5 comes from this
gene"
/codon_start=1
/protein_id="CAB03127.1"
/db_xref="GI:3877459"
/db_xref="SPTREMBL:P91997"
/translation="MSDSPANYPGCSQKNSGYNIPYNGFCYTKYTGENVPAIDAL
TAGYTRIPDAKYILNEKEGALKILLPEHGSKDSVEITSTSPESKAKRCACGV
SESLQSLQTDYIDMYLVHYKPRDSDNDPNNAEYKRLAYELAEKAGKRGVSGV
NNEIYHLEKELTYAKVPPCANOLEVHPHARIPLOKYEKNIFFQARSLSLRHPEKI
IEDPVVLEAKKHTNSVPLVLLAMLRQWGVIPKSPVPSRIVENKFKVIDALTPEDI
OSLPLADROGHVIRCTGMGVK"
18714..20251
/gene="F53F1.3"
CDs
join(18714..18887,18935..19056,19113..19302,19353..19448
19419..20008,20072..20251)
/gene="F53F1.3"
/note="similar to aldehyde reductase"
/codon_start=1
/protein_id="CAB03128.1"
/db_xref="GI:3877460"
/db_xref="SPTREMBL:P91998"
/translation="MSIKLNTGYDCPLGLGTYIKIGQVLPVLDAALTAGYRLPDTA
KYNNKEKEGDALEILPEHNLKREDIFPTTHMHPNTENVKIKLVDESILKTSYID
MYLHYKPSFGDGDODPMKTKLRIATWMDLMECKNAGIRSGVSGSEFRLHEELKD
GNPFPCCNOVEYVHPHFPREELKNYCKSGEIFQAFSSLRARNEFLSEITRLAEK
YHVPRTYVLSWATSQKYGIIIPKSTNPBELAQNLKTVLLEEEYVAKICNLNDKIYVR
TNGKVLV"
21582..22009
/gene="F53F1.4"
CDs
join(21582..21670,21781..22009)
/gene="F53F1.4"
/note="predicted using GeneFinder; similar to cuticlin;
cDNA EST EMBL:D72018 comes from this gene; cDNA EST
EMBL:D74799 comes from this gene; cDNA EST EMBL:C11739
comes from this gene; cDNA EST EMBL:C12451 comes from this
gene; cDNA EST EMBL:C10189 comes from this gene; cDNA EST
EMBL:C10682 comes from this gene; cDNA EST yk506b3.3 comes
from this gene; cDNA EST yk437e9.3 comes from this gene;
cDNA EST yk437e9.5 comes from this gene; cDNA EST
yk311h10.3 comes from this gene; cDNA EST yk311h10.5 comes
from this gene; cDNA EST yk341f6.3 comes from this gene;
cDNA EST yk341f6.5 comes from this gene; cDNA EST
yk300g2.3 comes from this gene; cDNA EST yk300g2.5 comes
from this gene; cDNA EST yk324b3.3 comes from this gene;
cDNA EST yk324b3.5 comes from this gene; cDNA EST
yk191d9.3 comes from this gene; cDNA EST yk191d9.5 comes
from this gene; cDNA EST yk677d4.3 comes from this gene"
/codon_start=1
/protein_id="CAB03125.1"
/db_xref="GI:3877457"
/translation="MOKIVIFPAALVAISOAFLPSSGGGGCGGCGAPPPPGCGGCS
LPLDLPRLPELRLSLSPFGGGGCGGAPPCGAAAYAYAPPAAGYAAPPAAGYATA
GRK"
25572..26408
/gene="F53F1.5"
CDs
join(25572..25660,25850..26408)
/gene="F53F1.5"

/note="pr:dictated using GeneFinder: similar to cuticlin";

cdna EST EMBL:J01107 comes from this gene; cdna EST
CESSN0F comes from this gene; cdna EST
cdna EST EMBL:D71580 comes from this gene;
cdna EST EMBL:D72120 comes from this gene; cdna EST
EMBL:D72428 comes from this gene; cdna EST
EMBL:D73526 comes from this gene; cdna EST EMBL:D74233 comes from this
gene; cdna EST EMBL:D74402 comes from this gene; cdna EST
EMBL:D74922 comes from this gene; cdna EST
EMBL:D75307 comes from this gene; cdna EST EMBL:D76322 comes from this
gene; cdna EST EMBL:C12136 comes from this gene; cdna EST
EMBL:D69807 comes from this gene; cdna EST
yk505b11.3 comes from this gene; cdna EST yk492b11.3 comes from this
gene; cdna EST yk492b11.5 comes from this gene; cdna EST
yk480b10.3 comes from this gene; cdna EST yk480b10.5 comes
from this gene; cdna EST yk445a7.3 comes from this gene;
cdna EST yk445a7.5 comes from this gene; cdna EST
yk440b8.3 comes from this gene; cdna EST yk440b8.5 comes
from this gene; cdna EST yk377b11.3 comes from this gene;
cdna EST yk377b11.5 comes from this gene; cdna EST
yk473a12.3 comes from this gene; cdna EST yk473g12.5 comes
from this gene; cdna EST yk459g4.3 comes from this gene;
cdna EST yk459g4.5 comes from this gene; cdna EST
yk482b10.3 comes from this gene; cdna EST yk482d10.5 comes
from this gene; cdna EST yk455c6.3 comes from this gene;
cdna EST yk455c6.5 comes from this gene; cdna EST
yk417a12.3 comes from this gene; cdna EST yk417a12.5 comes
from this gene; cdna EST yk414h1.3 comes from this gene;
cdna EST yk401f4.3 comes from this gene; cdna EST
yk401f4.5 comes from this gene; cdna EST yk401e5.3 comes
from this gene; cdna EST yk401e5.5 comes from this gene;
cdna EST yk399e6.3 comes from this gene; cdna EST
yk399e6.5 comes from this gene; cdna EST yk392h1.3 comes
from this gene; cdna EST yk392h1.5 comes from this gene;
cdna EST yk391b10.3 comes from this gene; cdna EST
yk391d10.5 comes from this gene; cdna EST yk342c6.3 comes
from this gene; cdna EST yk342c6.5 comes from this gene;
cdna EST yk345c4.3 comes from this gene; cdna EST
yk345d4.5 comes from this gene; cdna EST yk307f5.3 comes
from this gene; cdna EST yk307f5.5 comes from this gene;
cdna EST yk258a5.3 comes from this gene; cdna EST
EMBL:c00946 comes from this gene; cdna EST yk306e6.5 comes
from this gene; cdna EST yk314h1.5 comes from this gene;
cdna EST yk331e10.5 comes from this gene; cdna EST
yk343d4.5 comes from this gene; cdna EST yk375f1.5 comes
from this gene; cdna EST yk404c1.5 comes from this gene;
cdna EST yk447c9.5 comes from this gene; cdna EST
yk447e5.5 comes from this gene; cdna EST yk448e6.5 comes
from this gene; cdna EST yk382a5.3 comes from this gene;
cdna EST yk506h10.3 comes from this gene; cdna EST
yk507f12.3 comes from this gene; cdna EST yk518b6.3 comes
from this gene; cdna EST yk530f12.3 comes from this gene;
cdna EST yk536e1.3 comes from this gene; cdna EST
yk537e8.3 comes from this gene; cdna EST yk540g5.3 comes
from this gene; cdna EST yk545b10.3 comes from this gene;
cdna EST yk557b11.3 comes from this gene; cdna EST
yk562h11.3 comes from this gene; cdna EST yk564d11.3 comes
from this gene; cdna EST yk566f10.3 comes from this gene;
cdna EST yk596a10.3 comes from this gene; cdna EST
yk596e11.3 comes from this gene; cdna EST yk614f1.3 comes
from this gene; cdna EST yk618d8.3 comes from this gene;
cdna EST yk620c6.3 comes from this gene; cdna EST
yk620d5.3 comes from this gene; cdna EST yk622e4.3 comes

	Query Match	9.3%	Score 107.4	: DB: 34	Length 39478;
	Best Local Similarity	61.7%	Pred. No. 2.1e-15;		
	Matches 171; Conservative	0;	Mismatches 106;	Indels	0; Gaps 0;
QY	557 ccgtgataccttctgcgcgglttgccatctcccgatttcgatgatgtaacggtcata	616			
DB	15350 CAGGAGACTCATTTGCATGGTAGTTCACACAGTGCTTTCTGTGGACAGTGGACAGAGCTGAA	15291			
QY	617 ctgtgaacttcctaabctctgagattgcctctgataaaatatctgtctaataatttg	676			


```

Db 15290 CCCTCATTTCTCAATGACCTGCAATGGCTGCTCAATTGATGATTTCTATGTGCAACCTTG 15231
Oy 677 aatcacaacacattaatgctgccaagaagctcacgtatcacaaatgctgagctcgt 736
Db 15230 AGTATCCGGGAACCTTACTCGCTGGGCAAGACCTCATGTGTATTAAGTTTGTCGATCGAG 15111
Oy 737 caacagctttctatcaatgccaagatcagatattacattaaagaacaataagcaatg 796
Db 15170 ATGCTCTTTTTCACAGTGTGATATCAATTAACCGTCAAGACGACGACCAAGAAATGCG 15111
Oy 797 ttcgacacacatgctcagaaccagaagatctcgagc 833
Db 15110 TGAGACCAATATGTAAGATGTAGAAGGCGGTGAGC 15074

RESULT 10
CE004D5 31536 bp DNA INV 02-SEP-1999
LOCUS Caenorhabditis elegans cosmid E04D5, complete sequence.
DEFINITION 266496
ACCESSION 266496.1 GI:1041307
VERSION HTG: Cullicin-like protein.
KEYWORDS Caenorhabditis elegans.
SOURCE Caenorhabditis elegans.
ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilitida;
Rhabdilitae; Rhabdilitidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 31536)
Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kersey, J., Kirsten, J., Laister, N.,
Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A.,
Saunders, D., Showkeen, R., Smalton, N., Smith, A., Sonnenhammer, E.,
Steden, R., Sulston, J., Thierly-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Sproat, J. and Wohlman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
TITLE Nature 368 (6466), 32-38 (1994)
JOURNAL 94150718
MEDLINE 2 (bases 1 to 31536)
REFERENCE McMurray, A.
AUTHORS Direct Submission
JOURNAL Submitted (25-OCT-1995) Louis, MO 63110, USA. E-mail:
jes@anger.ac.uk or rwenemate@wustl.edu
COMMENT Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
For a graphical representation of this sequence and its analysis
see:-
http://webc.sanger.ac.uk/cgi-
bin/display?db=wormbase&class=Sequence &object=E04D5
Current sequencing finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone E04D5.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone E04D5 is at 1 in this sequence. The true
right end of clone E04D5 is at 2110 in
sequence 248585.
The true left end of clone E04D5 is at 31437 in this sequence. The
true right end of clone E04D5 is at 11412 in this sequence. The

```

```

FEATURES
    source
        start of this sequence (1..104) overlaps with the end of sequence
        249070.
        The end of this sequence (31437..31536) overlaps with the start of
        sequence 248585.
        location/Qualifiers
            1..31536
                /organism="Caenorhabditis elegans"
                /db_xref="taxon:6239"
                /clone="E04D5.1"
                /chromosome="II"
                /gene="E04D5"
                /gene="E04D5.1"
                /join(6082..6250,6307..6379,6432..6509,6564..6696,
                6750..6930,7003..7283,7337..8179,8228..8326)
                /gene="E04D5.1"
                /note="CDNA EST EMBL:D71941 comes from this gene; CDNA EST
                EMBL:D74691 comes from this gene; CDNA EST EMBL:D76330
                comes from this gene; CDNA EST EMBL:D65192 comes from this
                gene; CDNA EST EMBL:D68540 comes from this gene; CDNA EST
                EMBL:C11194 comes from this gene; CDNA EST EMBL:C13213
                comes from this gene; CDNA EST EMBL:C11881 comes from this
                gene; CDNA EST yk412d7.3 comes from this gene; CDNA EST
                yk412d7.5 comes from this gene; CDNA EST yk459d6.3 comes
                from this gene; CDNA EST yk459d6.5 comes from this gene;
                CDNA EST yk454d9.3 comes from this gene; CDNA EST
                yk454d9.5 comes from this gene; CDNA EST yk350d11.3 comes
                from this gene; CDNA EST yk350d11.5 comes from this gene;
                CDNA EST yk413f5.3 comes from this gene; CDNA EST
                yk413f5.5 comes from this gene; CDNA EST yk402d12.3 comes
                from this gene; CDNA EST yk402d12.5 comes from this gene;
                CDNA EST yk375c1.3 comes from this gene; CDNA EST
                yk375c1.5 comes from this gene; CDNA EST yk374f8.3 comes
                from this gene; CDNA EST yk374f8.5 comes from this gene;
                CDNA EST yk365e8.3 comes from this gene; CDNA EST
                yk365e8.5 comes from this gene; CDNA EST yk256f12.3 comes
                from this gene; CDNA EST yk256f12.5 comes from this gene;
                CDNA EST EMBL:C12803 comes from this gene; CDNA EST
                EMBL:C12996 comes from this gene; CDNA EST EMBL:C13136
                comes from this gene; CDNA EST yk268e5.5 comes from this
                gene; CDNA EST yk519e6.3 comes from this gene; CDNA EST
                yk533e1.3 comes from this gene; CDNA EST yk556e6.3 comes
                from this gene; CDNA EST yk578e6.3 comes from this gene;
                CDNA EST yk598a11.3 comes from this gene; CDNA EST
                yk640h2.3 comes from this gene; CDNA EST yk656h4.3 comes
                from this gene"
                /codon_start=1
                /protein_id="CAA91279.1"
                /db_xref="GI:3875451"
                /db_xref="SPTREMBL:Q19052"
                /translation="MPAPNTPPSRDSNSFVSSFSNFLPEKFTSPRLSOTLI
                QINKGDLVYAVRSESEGYLKRGLGDVAVFEQNTSRDPAVACNVAYSNNGOLFAP
                CNDQVRYVEFLATNKKEITCVLKRRTITSPKGVNTPPAVYVGRKTEKPKRP
                NVRYSLADGKHVSTFSAPKASNEPQESDDESLAARVSGSEVFTYTMSEDRDYLK
                VEGATNFALSPGPAPNVAIVPAVGSVPARVAVHVRSEFSPVAGNRKFSKAVM
                TNGRQGSLLILASVEDKTNQSYVGGDSLLINIQGESVAVLEKGPILYAKAMP
                NGRFPAVGYGMPAKVFNVRGVPIFDITLGNIDGSEVAVFAGNIVLIGCKNTAKK
                MEWDVETKEKILIEVNTKILFPMADPGDGEQCTAPRLIDNSRFMYHTRLMA
                ETHFESPEKLEWVRMPTGYNKAIRKIEFTDMAGLIPRKDDASHPLNNVAGAV
                KQACATYPPIHRLKRLGGSGSGSPSAAPFTGNGNORPAQFRANGNANPOPEPQOS
                EOEKKAFLKRVKVEIKLVKORVANGLQLOPOMEKIORENEYLSKLITI"
                /complement(8619..11718)
                /gene="E04D5.2"
                /complement(join(6619..8761,8813..8988,8995..10028,
                10089..10283,10330..10614,10730..10935,11582..11629,
                11691..11718))
                /gene="E04D5.2"
                /protein_id="CAA91282.1"
                /db_xref="GI:3875454"
                /db_xref="SPTREMBL:Q19055"
                /translation="MRCGENDFYADHISTIPKCKKLIRVFLSNACKHLFGFLKIP
                INFOLLAVERFRSLPSSIFLATVATCMULICUTYLLGVADAGIWRKNTLPLFLYR
                YIVVFIYAKVVOFAIPFILLITIFERYLMTCTERKRKRGITFOVLTLSFRKAFSAI

```


gene
CDS
/translation="MKSALIGEGRKKNOGLKPDPSKLANPSESKTFRSDPDIOFLTK
RKIKAVDQDSRLSLEKTIFFHSLSINLITLYLKALHFQTRPPFSKSPHENTLFTF
OLKMGALDSDSYGARGRLIKFSYIVTALISLIFESICIGYIMLARSQVALSDPS
LYVDGRLIVTILISLILNYLICYFAIFKEMRCQSVTCAVASIYAMLLIGGICIGN
FRDQYLTIRLIMKLTSLRELGYTHMKGTESMDLQSNFKCCGNGGNDIMOKYT
SKWVHOAPRLIPESCCIPSEIRERNSPDODAPPYITSCYEPLONDLLHVN
VASMCTINAIYOIVFYSFKTIKILNPLTKT"
complement(4270..15311)
/gene="F53B6.2"
complement(join(4270..4339,4493..4547,5071..5295,
5368..5662,5891..6148,6204..6752,6818..7287,7546..767
8104..8184,12576..12775,13246..13384,13931..14132,
14183..14294,14341..14457,14994..15155,15196)..15311))
/gene="F53B6.2"
/note="predicted using GeneFinder; similar to
thrombospondin like; cDNA EST EMBL:Z14404 comes from this
gene; cDNA EST EMBL:112571 comes from this gene; cDNA EST
yk195c9.3 comes from this gene; cDNA EST yk195c9.5 comes
from this gene; cDNA EST yk28b67.3 comes from this gene;
cDNA EST yk28b67.5 comes from this gene; cDNA EST
yk32ne3.5 comes from this gene; cDNA EST yk353h1.3 comes
from this gene"
/codon_start=1
/protein_id="CAB03121.1"
/db_xref="GI:4008366"
/translation="MKCSYTVFLFLFYLLIASFVDAISMAAMSPMSCTKTCGGGS
ROLRLRSLKSGSGSVRFKVCQAKTCEKSLRADITCGGEFLIDYRCGGCVRSRLT
GANFLMRPDDTFCOATSRVCSKSGSCIOYCGDLISSFREDYACVCGGRDTCN
GKFLMKVEEETACASNCDDIVDWSGAGRSIASTSOP1VVCVANA1TRGVPEYLCDK
LRPKLEAPRCPMLICPSRMMAADWTECPHCEGTRREYVCVOTAHNV1PDRFC
ENGTPRAEENECVSCSGRMWAGKSKCTACSCGGVRRHVCAGTSGDCEGAPRE
TTCYGIIC3IATNS1LMDNDRAVLDGNTFEGNMHNMOMAP1VACGMSVCSCTGG
VMSRFVECVAN1PSTISAP1KRLPMSECDODQBP1KJESCEGCLDPSK1SDBAPY
OMRYSDMTQCSASCLGKQKALK1CYVSTKSKVMSQCARRRPRPKSPCNDHP
PFWLSTK1SDSCSGSGTARRSVCAQYTSKSGVGA0AH1V1RDRCHEFKPDETEFC
NVVA0PA1WVSLSMKRHNK1K1MLK1ADWTECSGSDSEGRVLRDCE1RDSRGKTQ
RBPDECAKRPOTVEYCSGSPRELLSNRYFEQNA0K1KLTLG1GAVTL1YGT
SIKICPARKFKK1K1WKKNGK1K1ND4H1KVSANGL1VFP1HARMBDAGVYEC1PDR
LOGN1TAP1RDP1PASP1VDL1AP1R01P1STK1NRQ1R0VSKED1V1RE0AS1V1HNMN1SL
1EAL1TAP1NDEKAR1EQ1LNE1V1ARD1I0HMS1E0CKT1C1HVA0GT1AR1SK1VTFH
GE1IRVND1S1CE1AS1V1PPE1TR1PC1HRE1DC1RWMA1SQMS1C1R0CV1S1MA1K1RNV
1TCRF1NG1SV1D1I0H1CD1ITNR1BAT1MDC1PNO1MC1KEM1TS1PMG1SC1S1C1G1G1V0L1RL
1SCV1W1SS1R1PAC1R1CE0MR1R1BS1AR1CA1VD1P1PC1P1TAS1Y1QR1DAS1C0G1SR1CD
1IK1LH1SD1SL1EV1R0K1C1SC1TFCV1ERKKF"
complement(19892..20283)
/gene="F53B6.8"
complement(join(19892..20022,20241..20283))
/gene="F53B6.8"
/codon_start=1
/protein_id="CAB54255.1"
/db_xref="GI:5824510"
/translation="MDFEILVFLMAVAGTFAGISVFSHSHSEFPRHGLIGGGFNPY
GG1YOSNTWGRK"
complement(20516..21818)
/gene="F53B6.4"
complement(join(20516..20602,20648..20953,20999..21424,
21609..21818))
/gene="F53B6.4"
/note="predicted using GeneFinder; cDNA EST yk193h3.3
comes from this gene; cDNA EST yk193h3.5 comes from this
gene; cDNA EST yk411f5.3 comes from this gene; cDNA EST
yk411f5.5 comes from this gene"
/codon_start=1
/protein_id="CAB03117.1"
/db_xref="GI:3877449"
/translation="MIRISGOQPRFKSCDSDIDNKQFELYLKEARVCAS1ARTML
VPL1VIT1TSLPMVA1A1FCA1KNRRTYHAKKNNKSKSASKSTRGAS1GK1SR
TSKAS1HRSKSSK1SGTSGKSGKSGKSGKSGKSKVKTAT1TSGV0V1TAS
TGVSDKQNSN1SKSKSKSKSKSKRN1LBD1ADK1KKEKSGK1VAT1L1PK1T0OT1GS
OVGH1AE1VNS1K1SKEM1NAP1AK1LQ1Y01TG1VNO1LE1LNT1SERRA1V1IK1SD1SL
1RVN1SVY1FAP1R1RS1V1K1D1V1RL1NG1EQ1TK1DL1V1LL1TAN1G1ST1P1H1EAF1NQA1EH1RM
1NV1PL1V1AA1"

Query Match	Best Local Similarity	Score	DB	Length
Matches 127; Conservative	59.3%; Pred. No. 1e-07;	6.48;	74.8;	32412;
	0; Mismatches 87; Indels 0; Gaps 0;			
227	cactccattgattcatgaacgattgtgagcgtacacgattcttgaatccaaagtgtattc	286		
29831	CAATTCCATTGGACGACATATGCGGAGCAACACCTGCATCCAGTGGATAT	29772		
287	ttgtaacaacaactgtgtcatcttcgattcgaatcattatgtgtacaaagtgtatcgtg	346		
29771	CATTTTCTTCCATGACATGATGATCACTCGTTTCATCTTCTTTGTGTACGGGAATGATCGAG	29712		
347	catatcgatgaacatgcttttaccatgaaagctgtataaacaagttagtgcacagattgag	406		
29711	CATTGAGTTCGATGTTCTTCTCCGAGTCATTAAGATGATGAACCGGAATTTGATG	29652		
407	tatctgaatcaacaactgcttttcaaatcaat	440		
29651	TTGGGTATGTTAACTCTGAAGTATTAAATCAATT	29618		

RESULT 12
 LOCUS CER07E3/c
 DEFINITION Caenorhabditis elegans cosmid R07E3, complete sequence.
 ACCESSION Z49207
 VERSION 249207.1 GI:1067021
 KEYWORDS HTG; ADP, ATP carrier protein; Cuticulin; Cysteine proteinase.
 SOURCE Caenorhabditis elegans.
 ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 39736)
 AUTHORS Wilson, R., Alnsough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirkman, J., Lalster, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Ritken, A., Roopra, A., Saunders, D., Showkhen, R., Smaldon, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, D., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J. and Wohlman, P.
 TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
 JOURNAL Nature 368 (6466), 32-38 (1994)
 MEDLINE 94150718
 REFERENCE 2 (bases 1 to 39736)
 AUTHORS Cottage, A.
 JOURNAL Direct Submission
 COMMENT Submitted (05-MAY-1995) Louis, MO 63110, USA. E-mail: jess@sanger.ac.uk or twenematode.wustl.edu
 On Nov 21, 1995 this sequence version replaced gi:798823. Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information.
 For a graphical representation of this sequence and its analysis see:--
<http://webcage.sanger.ac.uk/cgi-bin/display?db=wormcage&class=Sequence&object=R07E3>
 Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones.
 EXCEPTIONS are indicated by an explicit note.
 IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
 This sequence is the entire insert of clone R07E3. The true right end of clone F41E7 is at 36774 in this sequence. The start of this sequence (1..115) overlaps with the end of sequence Z68106. The end of this sequence (39633..39736) overlaps with the start of sequence AL032624.
 FEATURES
 source
 location/Qualifiers
 1..39736
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /chromosome="X"
 /clone="R07E3"
 5080..8873
 /gene="R07E3.6"
 join(5080..5246,5409..5769,5826..5942,6381..6905,6955..7075,7125..7448,7495..7787,7831..7980,8037..8174,8222..8405,8470..8873)
 /gene="R07E3.6"
 /note="Serine and threonine rich protein. Carboxyl terminus is glycine rich. cDNA EST EMBL:CO8296 comes from this gene; cDNA EST YK259b4.5 comes from this gene; cDNA EST YK559c11.3 comes from this gene; cDNA EST YK574h3.3 comes from this gene; cDNA EST YK597h8.3 comes from this gene"
 gene
 /codon_start=1

/protein_id="CA89071.1"
 /db_xref="GI:3878959"
 /db_xref="SPTREMBL:O21811"
 /translation="MRVLCILVYVAFAASQPFSGMGNRENYVOFGYKDAISSMSD LKTIAMRNPKYNAPSRSFTIPPTFAEDLDNKKLRGGSITINKKATITLLPRY KGQNPFTTSTTTTSTTTTQKPTTSTTTTSTPTPTTTTTEATITLTTTLELPK QOLARLQVPMDSGEPSADGIDKKTKAKVFTDDNNVSOEADAKPSTHENG EHVSTPEESITENTTAEPASTASTKTKLSVKIRSEDEREDKLSDDKPYMKK LEEIKKKEELVERELAESVAIEPEKRVVTKRTPVPTSTTTEETTEATITV EETTTTKKAPPTTTTSTTTTTOPTTTTTEPTTTEPTTTEPTTEALENELEKAL EVM KSLDDEERIDEKNEVOKRKDERMAAKRIRNNPLKMKQKREKNNRHKTTTSLAT EOPKTTTIVASSTETSTVTDTEATTTTTPATTTTATTTTAKNSNIEADELEET AEKTTTKVTEGPTITTEPEYIKKNALVIGTKITLPALAEAINAKPIEASSASAP SSMTEAVVAASAAAPALPEIPAPQIALPOLPDQISOFPVVASIVDMPIILG SILOSSASARAAASVRYETKTPNIRIDIMADGSEMSLFGSGLREILNPGILK RDKLRIDAKMAKLEEAQRNNAVESAATCAPDLSGFPQYQOQPVATNGKLPYEP PPGYGVIPAPAPPPPTFRVAPVQVLPVQVOTQAPKPTLPPTLPVTTTKPTRTD EGYEGRGVKTTHFGDSGVMTGGSHSPTEPTYLNMPESEGLAPVAPAPPPAPATO ATAAVTPAKVPAAPSPFGGGGEGGGGGRNGRGGGSGSPVPEDEVERNSK TDFSEFSNNRKL"
 complement(11510..14654)
 /gene="R07E3.5"
 complement(join(11510..11773,12101..12476,12528..12827,13402..13542,13587..13696,13745..13863,13912..14026,14341..14654))
 /note="cDNA EST YK278a10.3 comes from this gene; cDNA EST YK278a10.5 comes from this gene; cDNA EST YK486f8.5 comes from this gene; cDNA EST YK603b8.3 comes from this gene"
 /codon_start=1
 /protein_id="CA89072.1"
 /db_xref="GI:3878960"
 /db_xref="SPTREMBL:Q21812"
 /translation="MIFLVFWFLILLIFLAVGTGTLGREYVNTLVYIPEKATLN EYEEEGSVDEMDPEFOHKKPRSSASQYKIRKSDIIDAKHETVAKOT TSVYVDDTLDPIAGMEAVIEDQVTFNSAOLPSNMLLSRTYSPFPEWOLITL IAGRFRTYVLPCHILALGAIYALMYSTITIGLVPAKRYKRLNRCMLMCRITIS RAFSSVIRFHDKENKANKGGICVANHTSPIDVWVLSQNCYAMTGQKAGFLQTT LRSFHHIHFPERGEAGDRKAYMDRNRHVDENKLPILIIPEEGTINNTSVMKRGS FEIGITPIYAVKTDRLTLDAFNMSAOSYGRYLSMSTSAIICDWWYLPAMRGEN EDSIFAKRVKRAIYAKKGLDLEMDGDKREKRSKLVYTOOKLYFERLARITTLNA MLEKSSDIILNMGISIEBERDELKQIDBDDDEEMRKISSLKPKKIGDHADH IKSD"
 complement(17917..19072)
 /gene="R07E3.4"
 complement(join(17917..18025,18074..18331,18378..18555,18604..18751,18800..18877,18947..19072))
 /gene="R07E3.4"
 /note="similar to ADP, ATP carrier protein"
 /codon_start=1
 /protein_id="CA89069.1"
 /db_xref="GI:3878957"
 /db_xref="SPTREMBL:O21809"
 /translation="WASISATLDVLETSKKCLASSAAAIKTTTAPDRKVLQLO ROSEFAMAENVNGIDCISKIRLEOGAAMLMGNGAGVARCIPNHTLWAEFDITRNL LKNVDRNDSFGKPLAGTFVSGGLGATTLGYPEDARPLADLVKRGKYGWY DCLKIKASEEVAWSYKGLSALQFVINSRAIPFGIDSIITSVEDERSLFAACMLI AOSITTSQWVCYPLDTPVRRSMQSGQIQVYSTDCWKTLVKKGGINFYGALIT NSLRSTGALITTFYERSKYN"
 complement(21213..23259)
 /gene="R07E3.3"
 complement(join(21213..21359,21654..21739,21787..21866,22011..22170,22220..22335,22428..22542,22637..22779,23254..23259))
 /gene="R07E3.3"
 /note="similar to cuticulin"
 /codon_start=1
 /protein_id="CA89068.1"
 /db_xref="GI:3878956"
 /db_xref="SPTREMBL:O21808"
 /translation="METSFGRLPSTNVVVAFNPEFLTKDRVKKVCCFYEMERRIO KVIQISMPPTMHSKQLNMPVKRYEVLDSGTGPVYFAVVGQVYHKWCDEHEMT FCMILWHSQFVDDGNGRQVOLLNDCGALDKYLTNLNEYPTDLMAGREAHVYKADRDN

MEDLINE	94150718
REFERENCE	2 (bases 1 to 36306)
AUTHORS	McMurray,A.
TITLE	Direct Submission
JOURNAL	Submitted (21-MAY-1996) Louis, MO 63110, USA. E-mail: jesus@sanger.ac.uk or rwenematecode.wustl.edu
COMMENT	<p>On Nov 4, 1996 this sequence version replaced gi:1370039. Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information.</p> <p>For a graphical representation of this sequence and its analysis see:--</p> <p>http://webace.sanger.ac.uk/cgi-bin/display?db=wormace&class=Sequence&object=M142</p> <p>Current sequence finishing criteria for the C. elegans genome unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.</p> <p>IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.</p> <p>This sequence is the entire insert of clone M142.</p> <p>The true right end of clone C4489 is at 5059 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence 273424.</p> <p>The end of this sequence (36203..36306) overlaps with the start of sequence 299276.</p>
FEATURES	Location/Qualifiers
SOURCE	1..36306
	/organism="Caenorhabditis elegans"
	/db_xref="taxon:6239"
	/chromosome="III"
	/clone="M142"
gene	3511..5795
	/gene="unc-119"
	join(3511..3570,4639..4750,4821..5201,5689..5795)
CDS	/gene="unc-119"
	/note="cDNA EST YK258a1.3 comes from this gene; cDNA EST YK258a1.5 comes from this gene; cDNA EST YK532a1.3 comes from this gene"
	/codon_start=1
	/protein_id="CAA97807.1"
	/db_xref="GI:3878736"
	/db_xref="SWISS-PROT:O10658"
	/translation="MKAPQOQSIAPGSATPPSPMPPTVFOAITTEALLAKNOI TPNVLAIPGITGGLCSPSANVYNIEETKQIDLDTEHYLFELAKPENTENLCA QAESARYVRKFAPNFLKLTGVATVEKVGDPITPHERMERHFFKDRLLCFDPER GFCHMPSRNHCEHIYEPQSLQOLMDMINPNBETRSDFYENKLVHMKADYSYD A"
gene	8540..12501
	/gene="M142.2"
	join(8540..8572,9335..9416,9468..9679,10373..10721,10772..10974,11033..11175,11871..12501)
CDS	/gene="M142.2"
	/note="predicted using GeneFinder; similar to cuticulin"
	/codon_start=1
	/protein_id="CAA97806.1"
	/db_xref="GI:3878735"
	/db_xref="SPTRMBL:Q21540"
	/translation="MRPIPYDISLSTISFLSLILICSNPIDNGLVSELIHECVTKH AVEVILLDASGSIQDDTFKKQLSFAMHLASRLNISEGSHMALIOYAEFKLESLG QFNPTOLEMIAORIEYOSGATNGOALRLTLEKLOGARPIKVAIVITDGGSDG VSEBSOLLRDADVVAIVAGYNLVNVLHOMTGNPNVRFVSESEFOLDRLASLTM SMCTERRPRGPEITICGPDRIQVAKASTKQPEGVAFVMDHYHDEECRAGPKFPDSS IGLTVPSACVHWHRSLNPKFGIVESVIVMFHSLMTKDTQVYKVOCFMEADKHV TGLVSVSMITVFEHQIYOMPOCAVTLRKAPDGIYRFATLGSVYHRCICEVEBA KDQFGMLVHSCYVDNGSDVILIDNSGCLDAVLSTPDYDTSRLATPYVAFYK ADRVLOFOCIITCLKYGDCGEGITPPONCKLRGDSGHHNHHEKRRKRLVRLAD GCVGTIDVTSVTVLEDRPACQOVYITTSFRSISTNOKIENS"
gene	17427..23053
	/gene="vab-7"

CDS	join(17427..17468,18628..18766,20758..20915,22487..22677,22840..23053)
	/gene="vab-7"
	/note="VAB-7 even-skipped homologue"
	/codon_start=1
	/protein_id="CAA97809.1"
	/db_xref="GI:3878738"
	/db_xref="SWISS-PROT:Q93899"
	/translation="MQSFIDSLIGVKNKVPGLVEMVAASRASSPPEEQOHHDPKGV VAAAAAAGRRHHHPYDRDGGOMRRYRTAPRSGLTGRLEEFKAKENYSKRTGELA AELNPECTIKVWFQNRMRMKDRORVGLGIANPPPOMAAYMLNFAVEMMKRTAASO FGATGPGNAGATGNNGSTSPSAASGLPPLPGFPSPFLSONSTRKSPSPSHDSSKSK NTSSDDDESKPVNSNPSPPSPYSTD"
gene	28250..33043
	/gene="M142.5"
	join(28250..28366,28422..28499,28558..28991,30672..30966,31500..31763,32933..33043)
CDS	/gene="M142.5"
	/note="predicted using GeneFinder; cDNA EST EMBL:T01774 comes from this gene; cDNA EST YK470a9.3 comes from this gene; cDNA EST YK470a9.5 comes from this gene; cDNA EST YK476b6.5 comes from this gene; cDNA EST YK563c12.3 comes from this gene; cDNA EST YK662c9.3 comes from this gene; cDNA EST YK679e8.3 comes from this gene"
	/codon_start=1
	/protein_id="CAA97808.1"
	/db_xref="GI:3878737"
	/db_xref="SPTRMBL:Q21541"
	/translation="MASCNLTASVAVIAGVETIETNGLSARGVISEFDTRKVLVD TKEMAIKPMIRINSEHLKIKVLSLESESEFAAKTEQFRONNPVNGTRTEEL DKTIGELTPNLMKSPALISIRGOQVYLKRIPTPCWCGEDIRVLGLVHKPEVSD VTKDAKATGPDSESAESAVLOVKILSPMTREYRAPLDFTGAIYANVSSSMST DDIPOLADITLALSMFOAEQOEKIEOLOEIIKIDEDMQLSQFVNDDETSKRLVE GVAALGSEAPRARGICVSPPLVFKFHEHEEKKQIOLTFEPDRGLKPTFE VHYVKKPDLPAELAKFDVITADPFLPAECIKRASHRLLRGKSDVVLCTGAI MEDYASRLMAHRTSEPRANNLANDPSCFANYQTITPC"
gene	join(34039..34101,34448..34750,35530..36306,299276.1:1:105..698,299276.1:1:364..1566,299276.1:1:1666..1772,299276.1:1:2495..2795,299276.1:3714..3928,299276.1:4056..4104,299276.1:5846..5958,299276.1:6478..6729,299276.1:7169..7338)
CDS	/gene="M142.6"
	join(34039..34101,34448..34750,35530..36306,299276.1:1:105..698,299276.1:1:364..1566,299276.1:1:1666..1772,299276.1:2495..2795,299276.1:3714..3928,299276.1:4056..4104,299276.1:5846..5958,299276.1:6478..6729,299276.1:7169..7338)
	/gene="M142.6"
	/note="similar to zinc finger, C3HC4 type (RING finger); cDNA EST EMBL:D67323 comes from this gene"
	/codon_start=1
	/protein_id="CAA97810.1"
	/db_xref="GI:3878739"
	/db_xref="SPTRMBL:O45962"
	/translation="MAPTGGGGQVQVLECCSICNRHNEFPLVSLICGVTRKCAE KPNQTPCPHDMKTKTHSPSEYNNVALLSVFPPKQCMTLSGAVSEAKRVDQISI QIAKFEEDASERGGTVSSREISRTLDKRYALLCYMRVEDGRLTKLKRGISEYV MIETLISQNTWHSOLMSAVRAGCOFALQAMODVRLILMTLETGCIRKLNLY MYVQVTLASDYPGVSKTGVHVVOLIRASCPGKSHSISLMOLKEERTYESLR EHDQIYQINAFESGLRIGPQMSULLYLAADSHSHMSITDKLQSKSYOQYBELRA LAGSQTMLVPARYELTQVYPCLEFFAGLEHEDTSKRMIGDALHQRITLLKHCSD DLRKMRKEERGVTILQAEVPGMGGGGGAGVAGISGLHPLYSDIDTGRSISHT NPKMNSHNSQTPPKQPRQRYOMGIRPNMGAGVSSDAPPIPSHQOQPPQFNSOHL PQRGGRGAGAPPPPOQMPMLIGDMGAPMMAQTEVLTDGGKVNTPORVYIM QSPHPLGGVAVMIPQOQMPPOQMPSPVCGPAGMPQMPPIPVQVPRPTMTATSP TGSVITYPAASPQOPHTITIQISIGVKKRSNFKLYRKISFLNPLDFLLIRKLK EKKAGDLEFEFKRISDTQFKRYPSSFSQDNNMPMDRSGGWAGTGMLLRSGADA EOLAKRYEELIKRLOPSEDDDDDEGIGHVSYVAGSYLDDRHHLPLMIVPTID LPALISFAMNPTEETMTIGEMQONRPAASLAPSSNOPMNVNASSATVOAECEN RKILDPERAKMTIMEKYSTCFHTLTKDWVAFYVLTNLFNFAVRRRRRAATTPO PVIPMOVYQVQPIVPAENENPNVPPPPPOQMPHVDASIGLTLPIRILVANHQN VVNSDKRIDVVERISEAQGNASEANHLRLRELRAVESOMAHLDPTYKNNCLRAL QOVDMELQOLHLPJVEG"

/note="predicted using GeneFinder: similar to collagen: CDNA EST EMBL:D69143 comes from this gene; CDNA EST EMBL:D65641 comes from this gene"
 /protein_id="CAB03513.1"
 /codon_start=1
 /db_xref="SPTREMBL:O94399"
 /translation="MTISGQLVAGTAAALCGVAIVPALEAFVLHIDINSFQSVED LAEFLEADMTTVVPREDVAVQPEFTLKFERRKYGAGVAGSAGSSGCGCPQK DCPAPPPPCGAPGDDGPAGRAGNGSGSTEDRMAADPNKYKCAAGPGSPGN GEPHVGAGVGDTGKDGCEPGDPEDEDEGPGGLGPGGDEGVPVGQGTGGQG PGVAPCAPGPGRGDGERGNGDGGGPGAPGADGCPGPDGSGDGVGAVGACADAYCCPPRSAMATGSSDSQSPATYEPAPRATGYSYRPAAPKGTDAAPTA PHPPAPVAPPKLHDYSPAPVADHAADAAAPQYKRRKRAAHMV"
 /gene="ceh-8"
 /complement(join(8777..8956,9538..9646,9702..9830,9879..9976,12270..12598,13623..13818))
 /gene="ceh-8"
 /note="predicted using GeneFinder: Similarity to Mouse homeobox protein 58 (TR:G51366)"
 /codon_start=1
 /protein_id="CAB03519.1"
 /db_xref="GI:3881555"
 /db_xref="SWISS-PROT:O94398"
 /translation="MSNPDTNNNSATSSITPNCCEFLMEVSTNLVIOFTYFVKLYPNY QIRAEFGGAPPPLPIEMNFKIKCSFAIIQLPKLSFEFSVASFSPYTPRPS FSFSPSDSSSSSSSKNRNDIKDKRKRRTPTTQLALAEAPFKHTYPPVYR ETIAKVLQPEPVQVYQNRNARKRREKDDCGGEEHSLKIDPMSPWSKSEKTT PPMLEPNTLSTHNNGISDEFTKSEKEYGCPFAATGTPADNTHSKTGANFHLN FESDSKSEKSEKSPOTPSSTSPFSEYHPPTIPIYIPNQSNAFMNYPMPPEYPIH EPQQLTHSOKNE"
 /complement(16560..21861)
 /gene="ZK265.5"
 /complement(join(16560..16667,16744..16929,16977..17133,17179..17280,17420..17619,17696..17763,17814..17848,18123..18173,19898..20006,20052..20188,20236..20358,20406..20562,20879..21047,21096..21217,21445..21489,21572..21710,21816..21861))
 /gene="ZK265.5"
 /note="CDNA EST yk481g5.5 comes from this gene; CDNA EST yk508e4.3 comes from this gene"
 /codon_start=1
 /protein_id="CAB03517.1"
 /db_xref="GI:3881553"
 /db_xref="SPTREMBL:O94403"
 /translation="MINATLEIRDHADDLTYWAMPIRNSHMPSSIGLYAVLF IISRTFLIGVSTAMRTQKFTLKLYVQKLFYTRIFHRQRNIFISLLMPELLVCY LLMARKVLALETRESIKCNDECHHEFGKSVAKNMFDSIFTAANRRHYMTVLTTPY GILTERIPATLVLYDEYKRSRWMLFGLFVQNTLFSCHMAVVTTSGLTQVLLISGVI FLSEFSAIVLVEFYENQORSLMTELDRHTNLTYSYOLKENLKTLYKARFESIS ILLIITAMIGNSLPTLNLNEDLITLVYVQVYHNSPRTVLTAFTEITNRYKRF NKARTFMRIRDSKDEMEKLELSDSRPSIIMSPRSSTRADSTPTTKRASNRP NNAIGTIPAIRAOLIFIDARKVALFYLAFTVYLSFESRIELDSYTYLYVQHSYLNQI GVKMGKQVTLVYGFVLFSSKAHNRDRDPIDVDCVLGTVGCWFSVQYKVLV LITSMGDKRRTLTQAQCSERKEGWTGPGYDISHCEFLMYSILIIIEEALVANHVOYR AIVHOMDGRREHDLRTLCIOEYFVAMFLHAFWFKOIIISVLYHIFIEELIGVAAV VCMFEYTMVLVAPGLASPIRRYGRK"
 /gene="ZK265.6"
 /complement(join(18275..18384,18431..18539,18656..18751,18874..19128))
 /gene="ZK265.6"
 /note="predicted using GeneFinder"
 /codon_start=1
 /protein_id="CAB03516.1"
 /db_xref="GI:3881552"
 /db_xref="SPTREMBL:O94402"
 /translation="MARDYKRGKPAAYTNRNRNOKYLLKKDNKKLSKSAVPIIKYTW DETKPRRNVNDGIAFNPDVAPJAEKRIIDAEPIDGVSIVPPEKKYVGRKN EKQAAHITSNLEOQYKEEELAPADGDRFLFHRHETELCYMLARHGDDPQAMNRDK NLMQTPYQNAKKITITHKESECKFLETA"
 /gene="ZK265.7"
 /complement(join(2756..31619))

CDS

join(27566..27635,27677..27835,28839..28913,29226..29303,29467..29611,29755..29902,29982..30024,30072..30293,30599..30709,30782..30854,30908,31418..31619)
/gene="ZK265.7"
/codon_start=1
/protein_id="CAB03518.1"
/db_xref="GI:3881554"
/db_xref="SPTREMBL:Q94404"
[translation="MRGSIILPLPKPAKMLISRLPFTRTSFLIIEPLHFKAKEM
SEIACFCIRRVILDKNPVGRGLPKRAKKKCYGLVEDCTTHAVGIRAKKHIE
WELSRVLDNRNLGVPEEYKSMTGKMPGSGIRCMCGGGENNDSETSLEYAF
TEGNDADAKISEMLKSEKDESKKFTTPEPLTEYTKRIITPKRTTLAKTTTT
TTSITSTITTEKNSFELYDALKLPVLSKTYITKFEPSIKQAVFISNIYRPRTA
TTTAATVPKSTKRIGTGTJAHPLPEPEENASAAVINIADINKATSLINLQCFVF
KGVLPNDISLDTFTEETRKLEKEMKEENMLADEENESNDISLDDQGI
TAEOREEEREERYRMERAAEERLQEDTILRIIEEERLEADKRPNRNREDDMG
SDENSTGSLILVSPFSAVILLIV"]

gene

join(33403..33438,34288..34402,34791..34918,34995..35304,35555..35799,35846..36069,375713.1:107..118',375713.1:188..290,375713.1:343..430,375713.1:495..556)

	Query Match	5.4%	Score 62.2	DB 34	Length 36069;
	Best Local Similarity	56.7%	Pred. No. 9.8e-05;		
	Matches 115; Conservative	0;	Mismatches 88;	Indels 0;	Gaps
OY	269	tgaatccacagctgtagtattctgttaacaacaacacgtctgcatttcgtttcatcattcaatttg	328		
Dd	34996	TGGATTCACAGTGGATGTACTGCTACCGGTATTCACAGTGCTGGTGTCTTTCCATCCCTTATTCA	35055		
OY	329	ttaccaagaattgatactgycatcatcgagtacaaactgcttttacatggaagctgtataaacag	388		
Dd	35056	TCACATAAAGCTTCATTAACAACATPTCATGTGCATTAATGCTTTTTGAAGAAGCAATAAAGGGC	35115		
OY	389	ttagtgcacagatlgagglatctlgaatcaacaacgcgtcttccaactcaaatltgccga	448		
Dd	35116	TGACGCTCGAGTTGGGTGTATTGATTTCCACACGACAGAAATTAGAACGACATCTGGA	35175		
OY	449	tgcagatatgcgcgttatgaaatt	471		
Dd	35176	TACCTGATGTCACATATTCAATT	35198		

Search completed: April 16, 2000, 04:55:03
Job time: 11112 sec

THIS PAGE BLANK (USPTO)

FT CDS 20..358
 FT /tag= a
 PN W09730070-A1.
 PD 21-AUG-1997.
 PD 19-FEB-1997; U02318.
 PR 20-FEB-1996; US-011868.
 PA (SMK) SMITHKLINE BEECHAM CORP.
 PI Black MT, Burnham MK, Hodgson JE, Knowles DUC, Nicholas RO,
 PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;
 DR WPI: 97-424969/39.
 DR P-PSDB: W27907.
 PT Novel polypeptide(s) from *Staphylococcus aureus* strain WCUH29 - used
 PT to isolate antimicrobial compounds, and in vaccines against *S.*
 PT *aureus* infection
 PS Claim 9: Pages 719-720; 989pp; English.
 CC The present sequence encodes a *Staphylococcus aureus* protein, that,
 CC based on homology with an *Escherichia coli* protein, is believed
 CC to be a N-acetylglucosamine-6-phosphate deacetylase. The present sequence
 CC was obtained from a library of clones of *S. aureus* WCUH 29 in *Escherichia*
 CC *coli*. The DNA sequence can be used in the construction of ribozymes and
 CC antisense sequences to control the expression of *Staphylococcus* genes.
 CC The DNA sequence is also useful as a source of regulatory elements for
 CC the control of bacterial gene expression. The encoded protein may be used
 CC to produce vaccines to enable a host to produce specific antibodies
 CC with antibacterial action. These vaccines and antibodies would protect
 CC a host against invasion by *S. aureus*, and conditions relating to
 CC *Staphylococcus* infection, e.g. *Staphylococcus* food poisoning, scaled
 CC skin syndrome, and toxic shock syndrome.
 SQ Sequence 541 BP; 180 A; 75 C; 107 G; 171 T;

Query Match 3.1%; Score 35.6; DB 1; Length 541;
 Best Local Similarity 52.5%; Pred. No. 0.83;
 Matches 74; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

OY 16 attgcttcgtacacattatgcatgtcttatttcacggttgcacatggtgc 75
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 89 ATTGCTTACCGTATGAAAGATGATGACGTNTTATTATTACCGATCAATGCGTGCA 148
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 76 gaagtgtacccagaatgaatggtgaccactcaataacatcaatttatacagt 135
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 149 AAGGATATGCCCGAAGAGAAATATGATNTGGGTGGACAAAGTAAGTTCATTCGCAA 208
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY 136 aatgcatcgaaagacatggt 156
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 209 CAAGCACGCTTGCAAAATGCT 229

RESULT 3
 X33788 X33788 standard; DNA; 348 BP.
 AC X33788;
 DT 25-JUN-1999 (first entry)
 DE *S. aureus* coding sequence SEQ ID NO. 45.
 KW upper respiratory tract infection; otitis media; bacterial tracheitis;
 KW acute epiglottitis; thyroditis; empyema; lung abscess; splenic abscess;
 KW cardiac infection; infective endocarditis; secretory diarrhea; ulcer;
 KW retroperitoneal abscess; cerebral abscess; blepharitis; conjunctivitis;
 KW keratitis; endophthalmitis; preseptal cellulitis; orbital cellulitis;
 KW dacryocystitis; epididymitis; intrarenal abscess; perinephric abscess;
 KW toxic shock syndrome; impetigo; folliculitis; cutaneous abscess;
 KW cellulitis; wound infection; bacterial myositis; septic arthritis;
 KW osteomyelitis; *Helicobacter pylori* infection; stomach cancer; gastritis;
 KW ss.
 OS *Staphylococcus aureus*.
 PN W09912557-A1.
 PD 18-MAR-1999.
 PR 14-SEP-1998; U18987.
 PR 12-SEP-1997; US-058710.
 PA (SMK) SMITHKLINE BEECHAM CORP.
 PI Burnham MKR, Lonetto MA, Warren PV;
 DR WPI: 99-229138/19.

DR P-PSDB: Y05315.
 PT New isolated *Staphylococcus aureus* polynucleotides
 PS Claim 21: Page 90-91; 102pp; English.
 CC This sequence represents a *S. aureus* polynucleotide of the invention.
 CC The invention also relates to the polypeptides encoded by the
 CC *S. aureus* polynucleotides.
 CC The polypeptides can be used for the treatment or prevention of disease.
 CC The polypeptide or polynucleotide can also be used to diagnose diseases
 CC related to their expression. The polypeptides and vectors containing them
 CC can also be used in immunisation methods. The products can be used for
 CC treating infection, e.g. infections of the upper respiratory tract,
 CC (e.g. otitis media, bacterial tracheitis, acute epiglottitis,
 CC thyroditis), respiratory (e.g. empyema, lung abscess), cardiac
 CC (e.g. infective endocarditis), gastrointestinal (e.g. secretory
 CC diarrhea, splenic abscess, retroperitoneal abscess), central nervous
 CC system (CNS) (e.g. cerebral abscess), eye (e.g. blepharitis,
 CC conjunctivitis, keratitis, endophthalmitis, preseptal and orbital
 CC cellulitis, dacryocystitis), kidney and urinary tract
 CC (e.g. epididymitis, intrarenal and perinephric abscess, toxic shock
 CC syndrome), skin (e.g. impetigo, folliculitis, cutaneous abscesses,
 CC cellulitis, wound infection, bacterial myositis), bone and joint
 CC (e.g. septic arthritis, osteomyelitis), or *Helicobacter pylori*
 CC infections, (e.g. causing stomach cancer, ulcers and gastritis). The
 CC products can also be used for treating in-dwelling devices and wounds.
 SQ Sequence 348 BP; 108 A; 55 C; 78 G; 107 T;

Query Match 3.0%; Score 35.4; DB 1; Length 348;
 Best Local Similarity 53.2%; Pred. No. 0.77;
 Matches 75; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

OY 16 attgcttcgtacacattatgcatgtcttatttcacggttgcacatggtgc 75
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 82 ATTGCTTACCGTATGAAAGATGATGACGTNTTATTATTACCGATCAATGCGTGCA 141
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 76 gaagtgtacccagaatgaatggtgaccactcaataacatcaatttatacagt 135
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 142 AAGGATATGCCCGAAGAGAAATATGATNTGGGTGGACAAAGTAAGTTCATTCGCAA 201
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 136 aatgcatcgaaagacatggt 156
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 202 CAAGCACGCTTGCAAAATGCT 222

RESULT 4
 X33781 X33781 standard; DNA; 1182 BP.
 AC X33781;
 DT 25-JUN-1999 (first entry)
 DE *S. aureus* coding sequence SEQ ID NO. 8.
 KW upper respiratory tract infection; otitis media; bacterial tracheitis;
 KW acute epiglottitis; thyroditis; empyema; lung abscess; splenic abscess;
 KW cardiac infection; infective endocarditis; secretory diarrhea; ulcer;
 KW retroperitoneal abscess; cerebral abscess; blepharitis; conjunctivitis;
 KW keratitis; endophthalmitis; preseptal cellulitis; orbital cellulitis;
 KW dacryocystitis; epididymitis; intrarenal abscess; perinephric abscess;
 KW toxic shock syndrome; impetigo; folliculitis; cutaneous abscess;
 KW cellulitis; wound infection; bacterial myositis; septic arthritis;
 KW osteomyelitis; *Helicobacter pylori* infection; stomach cancer; gastritis;
 KW ss.
 OS *Staphylococcus aureus*.
 PN W09912557-A1.
 PD 18-MAR-1999.
 PR 14-SEP-1998; U18987.
 PR 12-SEP-1997; US-058710.
 PA (SMK) SMITHKLINE BEECHAM CORP.
 PI Burnham MKR, Lonetto MA, Warren PV;
 DR WPI: 99-229138/19.
 DR P-PSDB: Y05308.
 PT New isolated *Staphylococcus aureus* polynucleotides
 PS Claim 20: Page 65-67; 102pp; English.
 CC This sequence represents a *S. aureus* polynucleotide of the invention.

CC The invention also relates to the polypeptides encoded by the
CC S. aureus polynucleotides.
CC The polypeptides can be used for the treatment or prevention of disease.
CC The polypeptide or polynucleotide can also be used to diagnose diseases
CC can also be used in immunisation methods. The products can be used for
CC treating infection, e.g. infections of the upper respiratory tract,
CC (e.g. otitis media, bacterial tracheitis, acute epiglottitis,
CC thyroditis), respiratory (e.g. empyema, lung abscess), cardiac
CC (e.g. infective endocarditis), gastrointestinal (e.g. secretory
CC diarrhoea, splenic abscess), retroperitoneal abscess), central nervous
CC system (CNS) (e.g. cerebral abscess), eye (e.g. blepharitis,
CC conjunctivitis, keratitis, endophthalmitis, preseptal and orbital
CC cellulitis, dacryocystitis), kidney and urinary tract
CC (e.g. epididymitis, intrarenal and perinephric abscess, toxic shock
CC syndrome), skin (e.g. impetigo, folliculitis, cutaneous abscesses,
CC cellulitis, wound infection, bacterial myositis), bone and joint
CC infections, (e.g. causing stomach cancer, ulcers and gastritis). The
CC products can also be used for treating in-dwelling devices and wounds.
SQ Sequence 1182 BP; 416 A; 174 C; 257 G; 335 T;

Query Match 3.0%; Score 35.4; DB 1; Length 1182;
Best Local Similarity 53.2%; Pred. No. 1.4;
Matches 75; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 16 attgcttcctgactacattgcatgtcttattcgatccggttgacaatggtctc 75
DB ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
805 ATTGCTTACCGTATGAAGGTATGACGCTTTTATTATTAATACGATGCGATGCGCA 864
QY 76 gaaggtgagccagaatgaatgacgaacttcaatacatatattatcaagct 135
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
865 AAGGTATGCGCTGAAGGAGAAATATGATTTGGGTGACAAAAGTACTGTTCAATCGCA 924
QY 136 aatgcattcgaaagacatggt 156
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
925 CAAAGCACGCTTGCACAAATGCT 945

RESULT 5
V74344 ID V74344 standard; DNA: 7563 BP.
AC V74344;
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #33.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT misc-feature 481..540
FT /*tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc-feature 2281..2340
FT /*tag= b
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc-feature 4081..4140
FT /*tag= c
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc-feature 5881..5940
FT /*tag= d
FT /note= "these bases represent a line of missing text in

FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT

EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA,
DR WPI: 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1: Page 321-325; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC and their fragments are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 7563 BP; 2584 A; 1028 C; 1476 G; 2229 T;

Query Match 3.0%; Score 35.4; DB 1; Length 7563;
Best Local Similarity 53.2%; Pred. No. 3.2;
Matches 75; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 16 attgcttcctgactacattgcatgtcttattcgatccggttgacaatggtctc 75
DB ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5476 ATTGCTTACCGTATGAAGGTATGACGCTTTTATTATTAATACGATGCGATGCGCA 5535
QY 76 gaaggtgagccagaatgaatgacgaacttcaatacatatattatcaagct 135
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5536 AAGGTATGCGCTGAAGGAGAAATATGATTTGGGTGACAAAAGTACTGTTCAATCGCA 5595
QY 136 aatgcattcgaaagacatggt 156
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5596 CAAAGCACGCTTGCACAAATGCT 5616

RESULT 6
V35363/C ID V35363 standard; DNA: 4940 BP.
AC V35363;
DT 23-SEP-1998 (first entry)
DE P. falciparum gp190 DNA.
KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
KW monoclonal antibody; passive immunisation; parasite; ss.
OS Plasmodium falciparum.
FH Key Location/Qualifiers
FT CDS 10..4929
FT /*tag= a
FT /product= gp190
FT
FT WO9814583-A2.
PD 09-APR-1998.
PF 02-OCT-1997; E05441.
PR 02-OCT-1996; DE-040817.
PA (BUOA/) BUOARD H.
PI Bujard H, Pan W, Tolle R;
DR WPI: 98-240088/21.

Dd	4247	GATGCCCTATTTCCTGTTGTAGAACACATTAAACAATGCTGGTGGTTGATTTCATTCCAGAT	4188
Oy	661	ttgtcctaataatttgyaatatccaaagatitaatgctggccaagaagctcaagtlac	720
Dd	4187	ATCGAATAATTGGGTGGATTGCACAGAATAACTGCTGAANAATGTGGCTCAAGAANTT	4128
Oy	721	aaatatcgagatcg	734
Dd	4127	GGTTACGACAGATGG	4114
 RESULT 9 X13342			
ID	X13342	standard; DNA: 1479 BP.	
AC	X13342;		
DT	19-MAR-1999	(first entry)	
DE	Enterococcus faecalis genome contig SEQ ID NO:405.		
KW	Enterococcus faecalis; config; detection; Enterococcal infection;		
KW	vaccine; attenuation; computer readable medium; ds.		
OS	Enterococcus faecalis.		
PN	MO9850555-A2.		
PD	12-NOV-1998.		
PE	04-MAY-1998; U08985.		
PR	14-NOV-1997; US-066009.		
PR	06-MAY-1997; US-044031.		
PR	16-MAY-1997; US-046655.		
PA	(HUMA-) HUMAN GENOME SCI INC.		
PI	Barash SC, Dillon PJ, Kunsch CA;		
DR	WPI: 99-045171/04.		
PT	New isolated Enterococcus faecalis polynucleotides and polypeptides		
PT	- used to develop products for the detection of Enterococcus and for		
PT	use in vaccines for prevention or attenuation of Enterococcus		
PT	infection.		
PS	Claim 1: Page 1617-1618; 2084pp; English.		
CC	A computer readable medium has been developed which has recorded on it		
CC	982 nucleotide sequences isolated from the Enterococcus faecalis genome,		
CC	X12938 to X13919 represent these nucleotide sequences which are primary		
CC	nucleotide sequences, also known as configs. The computer-based system		
CC	can identify fragments of the Enterococcus faecalis genome with		
CC	commercial importance. The products can be used to detect the presence		
CC	of Enterococcus faecalis in samples. They can also be used for		
CC	diagnosing Enterococcal infection in an animal and monitoring		
CC	progression of disease, and for identifying agents which can be used to		
CC	modulate the growth or pathogenicity of Enterococcus faecalis, or		
CC	another related organism, in vivo or in vitro. In particular the		
CC	polypeptides encoded by the Enterococcus faecalis nucleotide sequences		
CC	can be used in vaccines to prevent or attenuate an Enterococcal		
CC	Infection.		
SQ	Sequence 1479 BP; 521 A; 236 C; 316 G; 399 T;		
 Query Match 3.0%; Score 34.6; DB 1; Length 1479; Best Local Similarity 44.3%; Pred. No.2.5; Matches 136; Conservative 1; Mismatches 170; Indels 0; Gaps 0;			
Oy	111	aataacaatcaattatatacacgaatgcattcgaagacagtgttatgtgaagguct	170
Dd	1061	AAGAACTATTATTATTAATGAATTTAGCAGATATGAAAATGCATCATGCAAGGAATGT	1120
Oy	171	ttaigtacaagaagcttgccgtaatgatgaagtgtagcgtcaagttgccgaatttcact	230
Dd	1121	TGAAAGTCAGAGTCGTCGTTATTTATGTAACCTTAGGTAAATCGAAGCAGCTTATTC	1180
Oy	231	tcacattgatcatcatcagatgltggcgcgaacagatcctctgaatccacgttgtatttgt	290
Dd	1181	AAAAAAGATCAATATGCCAAATGAATTTTACCACCCGCAATGRTCGGRITTAAGATTACGT	1240
Oy	291	aacaacaactgtgtcattcgtttccatccattatttgttaccaaagttagtcgtgcata	350
Dd	1241	ATCACGTGTGTAAGAAACATTTCAAAGGCCCAAGATTTTTTGTAAGCCGTAGTCATCCAGA	1300
Oy	351	tcgagtlacatgcttttacatgtgaagctgataaaacagtltagtcacagattgagtlac	410

	DB	1301	TTTACTAAACCKCTTGTGGAACAAGAAATTCAGAGACTTTATGCATTGGTAATAA	1360
OY	411	tgaatc 417 I I I		
DB	1361	CCTAAC 1367		
RESULT	10			
ID	V23659			
AC	V23659 standard; cDNA; 2155 BP.			
DN	17-AUG-1998 (first entry)			
DE	Human interleukin-1 receptor accessory molecule cDNA.			
KW	Interleukin-1 receptor accessory molecule; IL-1R ACM; human;			
RK	signal transduction; infection; septic shock; inflammation;			
OS	rheumatoid arthritis; therapy; ds.			
FH	Homo sapiens.			
FT	Key	Location/Qualifiers		
FT	CDS	303..1373	/tag= a	
FT	sigs.peptide	303..353	/tag= b	
FT	mat.peptide	354..1370	/tag= C	
PX	M09808969-A1.			
PD	05-MAR-1998.			
PR	26-AUG-1996: U13954.			
PA	(HUMAN) HUMAN GENOME SCI INC. Bednarik DP, Olsen HS, Rosen CA; P-PADB: M53897.			
PT	Nucleic acid encoding interleukin-1 receptor accessory protein - used for therapeutic modulation of IL-1 activity			
PS	Claim 2; Fig 1; 95pp: English.			
CC	This cDNA clone codes for human interleukin-1 receptor accessory molecule (IL1-R ACm) (see W53697), a new member of the immunoglobulin superfamily that forms a complex with type 1 IL1-R and which has higher affinity for IL-1 than the receptor itself, suggesting that the known high and low affinity forms of IL1-R are in fact the receptor with or without IL1-R ACm, respectively.			
CC	The 2155 bp sequence is present in clone hMEB52 (deposited as ATCC 97666) derived from microvascular epithelium (no details of isolation given). Recombinant expression in Escherichia coli,			
CC	mammalian and insect cells is described. Recombinant host cells and recombinant vectors are claimed. Also claimed are isolated nucleic acid molecules encoding epitope-bearing portions (see W53898-915) of IL1-R ACm. Recombinant IL1-R ACm can be used to identify IL-1R agonists and antagonists useful for therapeutic modulation of IL-1 activity, and to raise specific antibodies.			
CC	Nucleic acid fragments are useful as diagnostic probes and primers, for isolation of IL1-R ACm-encoding genomic sequences, for in situ hybridisation to determine chromosomal localisation of the gene, CC for chromosome identification and for identifying mutations that might be linked to disease.			
SQ	Sequence 2155 BP:	709 A:	429 C:	433 G: 584 T:
	Query Match	3.0%;	Score 34.6;	DB 1: Length 2155;
	Best Local Similarity	49.7%;	Pred.No. 3;	
	Matches 88;	Conservative 0;	Mismatches 89;	Indels 0; Gaps 0;
OY	615	tacttctgaaattccaataagtcgtatggatgagtgccttttgataaatatttgtccaaaattaatt	674	
DB	1798	TCTCCTCAACAAAATMACAGAAGGAAAGTAATTAACCCTTTAAATTAATTAAGAACCATTAAC	1857	
OY	675	ggaacatccaagaatctaattgatcgcygcgcacaagaatcaaacatatcataaatatgcgatcg	734	
DB	1858	GTAGTGTTGNRAACATTATTTATATAGCACGCCATCCCAATTGATGCAAACTAATTAAGSTATTG	1917	
OY	735	alcacagcttcttaaccaatgcccatcatgatlaccatlaagaagacaatalagcga	791	
DB	1918	AATGTTTTATTCCCCAAAAATCGATAATTATATATTATTTTAAACATATGATCAAC	1974	

RESULT	11
V74396/c	
ID	V74396 standard; DNA; 9062 BP.
AC	V74396;
DT	16-MAR-1999 (first entry)
DE	Staphylococcus aureus contig SEQ ID #85.
KM	Computer readable medium; vaccine; S.aureus infection; immunodetection;
KM	cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KM	skin infection; surgical wound infection; scalded skin syndrome;
KW	toxic shock syndrome; ds.
OS	Staphylococcus aureus.
FH	Key
FT	Location/Qualifiers
FT	misc_feature
FT	721..780
FT	/tag-a
FT	/note-
FT	"these bases represent a line of missing text in
FT	the sequence listing in the specification. They
FT	are included to maintain the nucleotide numbering
FT	given in the specification for this DNA sequence"
FT	2521..2580
FT	misc_feature
FT	/tag-b
FT	/note-
FT	"these bases represent a line of missing text in
FT	the sequence listing in the specification. They
FT	are included to maintain the nucleotide numbering
FT	given in the specification for this DNA sequence"
FT	4321..4380
FT	misc_feature
FT	/tag-c
FT	/note-
FT	"these bases represent a line of missing text in
FT	the sequence listing in the specification. They
FT	are included to maintain the nucleotide numbering
FT	given in the specification for this DNA sequence"
FT	6121..6180
FT	misc_feature
FT	/tag-d
FT	/note-
FT	"these bases represent a line of missing text in
FT	the sequence listing in the specification. They
FT	are included to maintain the nucleotide numbering
FT	given in the specification for this DNA sequence"
FT	7921..7980
FT	misc_feature
FT	/tag-e
FT	/note-
FT	"these bases represent a line of missing text in
FT	the sequence listing in the specification. They
FT	are included to maintain the nucleotide numbering
FT	given in the specification for this DNA sequence"
PN	EP-786519-A2.
PD	30-JUL-1997.
PF	07-JAN-1997; 100117.
PR	05-JAN-1996; US-009861.
PA	(HUMA-) HUMAN GENOME SCI INC.
PI	Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI	Rosen CA;
DR	WP1: 97-374922/35.
PT	Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT	stored on computer readable medium and used in the production of
PT	anti-S.aureus vaccines
CT	Claim 1: Page 544-549; 3271pp; English.
CC	This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC	of the invention. The DNA sequences are recorded on a computer readable
CC	medium, preferably selected from a floppy or hard disk, random access
CC	memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC	the S.aureus DNA sequences allows putative functions to be assigned so
CC	that protein-encoding or regulatory regions of commercial, therapeutic or
CC	industrial importance can be obtained. Specifically, sequences which are
CC	likely to encode antigens have been identified and these polypeptides can
CC	be used in a vaccine composition against S.aureus infection. The
CC	S.aureus can also be used in a kit for the immunodetection of
CC	S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC	including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC	skin and surgical wound infections, scalded skin syndrome, toxic shock
CC	syndrome, etc. Organisms transformed with the DNA sequences can be used
CC	for recombinant production of the polypeptides. The new DNA sequences
CC	(and their fragments) are useful as primers or probes for isolating
CC	homologues of any of the S.aureus DNA sequences confined on the

CC	computer	readable	medium	Sequence	9062 BP	2756 A	1655 C	1271 G	3076 T
Query Match	3.0%	Score 34.6	DB 1	Length 9062					
Best Local Similarity	37.2%	Pred. No. 5.8							
Matches 94	Conservative 0	Mismatches 159	Indels 0	Gaps 0					
QY	91	attgaatgtgagccacttcaataaacaalcaatltthaataccgtaatgcatcgaagga	150						
DB	2747	ATTGGATTTCACAACAACACTGCTGTGTAATTCGAATTTGATGACGGTTATGATGGCAATCA	2688						
QY	151	catgtttatgtctaaaggtcttataatcatcaagaaggttcgcgaatgataagatgtgagct	210						
DB	2687	CTTTGTTAATGCAATATGTGTTGTTATCAATCAACGACATGATTCATAAAGCACAGCA	2628						
QY	211	caagtgcgcggaattccattccattgtaattcatgcaatgtgcgtaacagatctctg	270						
DB	2627	AAAGGTGTATGGTAATTCGCTCATTTATGTTGGTTTGAAACACTGGTCGNNNNNNNNNN	2568						
QY	271	aatccacgtgtgattcttgttaacaaacacgtgtgcatcattgcttccatcatattgtt	330						
DB	2567	NNNACGACTTCTGTGA	2508						
QY	331	accaaagtgtac	343						
DB	2507	CAAAATCGGTGATC	2495						
RESULT 12									
ID	X13031/c								
AC	X13031 standard; DNA; 21252 BP.								
DT	X13031.								
DE	19-MAR-1999 (first entry)								
KE	Enterococcus faecalis genome config SEQ ID NO:94.								
KM	Enterococcus faecalis; config; detection: Enterococcal infection;								
OS	vaccine; attenuation; computer readable medium; ds.								
PN	Enterococcus faecalis.								
PD	WO9850555-A2.								
PF	12-NOV-1998.								
PR	04-MAY-1998: U08985.								
PR	14-NOV-1997: US-066009.								
PR	06-MAY-1997: US-044031.								
PA	16-MAY-1997: US-046655.								
PI	(HUMA-) HUMAN GENOME SCI INC.								
PI	Barash SC, Dillon PJ, Kunsch CA;								
DR	WPI: 99-045171/04.								
PT	New isolated Enterococcus faecalis polynucleotides and polypeptides								
PT	- used to develop products for the detection of Enterococcus and for								
PT	use in vaccines for prevention or attenuation of Enterococcus								
PT	infection.								
PS	Claim 1: Page 615-626: 2084pp: English.								
CC	A computer readable medium has been developed which has recorded on it								
CC	982 nucleotide sequences isolated from the Enterococcus faecalis genome.								
CC	X12938 to X13919 represent these nucleotide sequences which are primary								
CC	nucleotide sequences, also known as configs. The computer-based system								
CC	can identify fragments of the Enterococcus faecalis genome with								
CC	commercial importance. The products can be used to detect the presence								
CC	of Enterococcus faecalis in samples. They can also be used for								
CC	diagnosing Enterococcal infection in an animal and monitoring								
CC	progression of disease, and for identifying agents which can be used to								
CC	modulate the growth or pathogenicity of Enterococcus faecalis, or								
CC	another related organism, in vivo or in vitro. In particular, the								
CC	polypeptides encoded by the Enterococcus faecalis nucleotide sequences								
CC	can be used in vaccines to prevent or attenuate an Enterococcal								
CC	infection.								
SEQ	Sequence	21252 BP;	6141 A;	4384 C;	3639 G;	7073 T;			
Query Match	3.0%	Score 34.6	DB 1	Length 21252					
Best Local Similarity	51.6%	Pred. No. 8.6							
Matches 79	Conservative 0	Mismatches 74	Indels 0	Gaps 0					


```
Oy 263 attcttgtaacacacacgctgttcattctgcgtttccatcatattgtttaccacaagtgtat 342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1701 ATTGCTGGGATTTTATCTGATTAATGTTGGTGGCGGTAATTTTGTCTTTTGGC 1642
Oy 343 cgtgcataatgagtcacatgctttacatggaagctgtataaacagcttagtgcacagatt 402
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1641 AAAAGCCAGTTGGAAGAACTCTGTCACACATCACTGAAACACAGTGAAATCATCACT 1582
Oy 403 gaggtatctgaatacacacactgctttccaact 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1581 GAAATTTCTGAAAGTAGTACAGTTTCTTCACT 1549

RESULT 13
X20248_02/c
Continuation (3 of 10) of X20248 from base 200001 (Borrelia burgdorferi polynucleotide s
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248
WP Fragment Name Begin End
WP X20248_00 1 110000
WP X20248_01 100001 210000
WP X20248_02 200001 310000
WP X20248_03 300001 410000
WP X20248_04 400001 510000
WP X20248_05 500001 610000
WP X20248_06 600001 710000
WP X20248_07 700001 810000
WP X20248_08 800001 910000
WP X20248_09 900001 910715
```

```
Query Match 3.0%; Score 34.6; DB 1; Length 110000;
Best Local Similarity 48.3%; Pred. No. 18;
Matches 97; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

Oy 529 taccataatgagcactgcatctcgtgaacgcgttgatactttctgcggcgtgtccattcc 588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 90819 TATCAAAATCAAAAAGCGATCTTTAGTGTGAATCTCCCTTAGAGGTGAATATCT 90760
Oy 589 tgccttgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcat 648
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 90759 TCATCTATTATTCTTATATTAAGTTCAGTAAGTATCTCTCTTAAGGTAATGAA 90700
Oy 649 ctgtataaatattgctaataattggaatataccaacagcttaagtgcgtgcagaaga 708
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 90639 TCTTTTGAGAAAGATTAATTTATCTGAAAGATTGAATGATATTTCGAACAATAAGAA 90640
Oy 709 gctcacgtataacaatatgcg 729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 90639 ACTGCTATATCAAAATACGAG 90619
```

```
RESULT 14
X13796/c
ID X13796 standard; DNA; 3549 BP.
AC X13796;
DT 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:859.
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.
PN W09850555-A2.
PD 12-NOV-1998.
PE 04-MAY-1998; U08985.
PR 16-MAY-1997; U08985.
PR 06-MAY-1997; U08985.
PR 16-MAY-1997; U08985.
PA (HUMAN-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PJ, Kunsch CA;
PI WPI: 99-045171/04.
DR New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
```

```
PS Claim 1; Page 2022-2024; 2084p; English.
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC X12938 to X13919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
SQ Sequence 3549 BP; 1435 A; 641 C; 682 G; 788 T;
```

```
Query Match 2.9%; Score 34.2; DB 1; Length 3549;
Best Local Similarity 52.4%; Pred. No. 4.9;
Matches 75; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Oy 285 ttgttaacacacacgctgttcattctgcgtttccatcatattgtttaccacaagtgtatcgc 344
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 314 TTTTCTAAACAGCAGCTGCTGATTAATCAATGATGATGATGATGATGATGATGATGAT 255
Oy 345 tgcataatgagtcacatgctttacatggaagctgtataaacagcttagtgcacagattga 404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 254 TGTGTTGGAAGCTTTTTCGCGTTCAGTTGTCGATCTTCTTCATATTAATTTGT 195
Oy 405 ggtatctgaatacacacactgctt 427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 194 CGTACTTTTAGTCGCACTGCTT 172
```

```
RESULT 15
T36122
ID T36122 standard; DNA; 4865 BP.
AC T36122;
DT 26-OCT-1996 (first entry)
DE S. mutans antigen I/II gene.
KW Caries; antigen I/II; epitope; tooth decay; vaccine;
KW genetic immunisation; ss.
OS Streptococcus mutans.
FH Key location/Qualifiers
FT CDS 100..4785
FT /tag= a
FT /transl_except= 2001..2049
FT /note= "bases 2001-2049 code for an amino acid
FT sequence (vgvarkqlvryfyfkskmkrlklsiqwl)
FT that differs from amino acids 618-650
FT given on page 46 of the specification"

W09623886-A1.
PD 08-AUG-1996.
PF 31-JAN-1996; G00207.
PR 31-JAN-1995; GB-001826.
PA (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PI Kelly C, Lehner T;
PI WPI: 96-371434/37.
DR P-PsDB; W02098.
DR Nucleic acid encoding polypeptide for prevention or treatment of
PT dental caries - which stimulates T or B cell response, and/or
PT adheres to tooth in competition with Streptococcus mutans antigen
PT I/II
PS Disclosure: Page 46-49; 63pp; English.
CC A Streptococcus mutans gene (T36122) codes for I/II antigen
CC (W02098), a 185 kDa cell surface protein at least partly responsible
CC for S. mutans adhesion to teeth. The I/II antigen includes a
CC series of overlapping T-cell, B-cell and adhesion epitopes.
CC Fragments (see also T36111-21) of the gene can be used to produce
CC recombinant polypeptides (W02087-97) carrying such epitopes for
CC use in vaccines for immunisation against dental caries. The DNA
CC fragments can also themselves be used as naked nucleic acid
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 16, 2000, 04:36:19 ; Search time 165.68 Seconds
(without alignments)
838.958 Million cell updates/sec

Title: US-09-323-427-3
Perfect score: 1161
Sequence: 1 atgatgatctcttcttc.....ttcttccaatcagaagca 1161

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 214294 seqs, 59861574 residues

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS9.COMB.seq:*
7: /cgn2_6/ptodata/1/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33.4	2.9	1867	1 US-07-772-087-1	Sequence 1, Appli
2	33	2.8	11225	7 5182210-9	Patent No. 5182210
3	32.6	2.8	5177	7 5352450-1	Patent No. 5352450
4	32.4	2.8	576	6 PCT-US96-05320A-705	Sequence 705, App
5	32.4	2.8	4090	4 US-08-781-802-5	Sequence 5, Appli
6	32	2.8	2338	4 US-08-425-069-1	Sequence 1, Appli
7	32	2.8	2338	4 US-08-317-844B-1	Sequence 1, Appli
8	31.6	2.7	4437	2 US-08-559-303B-72	Sequence 72, Appli
9	31.4	2.7	1189	6 PCT-US92-08328-4	Sequence 4, Appli
10	31.4	2.7	3414	2 US-08-764-100-17	Sequence 17, Appli
11	31.4	2.7	3414	2 US-08-764-100-21	Sequence 21, Appli
12	31.4	2.7	4970	2 US-08-764-100-14	Sequence 14, Appli
13	31.4	2.7	4970	2 US-08-764-100-20	Sequence 20, Appli
14	31.4	2.7	5655	4 US-08-989-478-1	Sequence 1, Appli
15	31.4	2.7	7295	4 US-08-487-826B-15	Sequence 15, Appli
16	31.2	2.7	14602	2 US-08-597-236-1	Sequence 1, Appli
17	31.2	2.7	14602	2 US-08-746-682A-1	Sequence 1, Appli
18	31.2	2.7	8855	3 US-08-542-003-1	Sequence 1, Appli
19	31	2.7	8855	3 US-08-322-760A-1	Sequence 1, Appli
20	30.8	2.7	325	1 US-08-111-316-6	Sequence 6, Appli
21	30.8	2.7	325	1 US-08-468-405-6	Sequence 6, Appli
22	30.8	2.7	326	1 US-08-111-316-7	Sequence 7, Appli
23	30.8	2.7	326	1 US-08-468-405-7	Sequence 1, Appli
24	30.8	2.7	1539	3 US-08-828-596-1	Sequence 1, Appli
25	30.8	2.7	2254	3 US-08-635-066-1	Sequence 1, Appli
26	30.8	2.7	2254	3 US-08-635-066-1	Sequence 1, Appli
27	30.6	2.6	5049	2 US-08-336-345-1	Sequence 1, Appli

28	30.6	2.6	5049	2 US-08-336-345-2	Sequence 2, Appli
29	30.6	2.6	5049	3 US-08-647-655-1	Sequence 1, Appli
30	30.6	2.6	5049	3 US-08-647-655-2	Sequence 2, Appli
31	30.6	2.6	19124	4 US-08-487-826B-13	Sequence 13, Appli
32	30.2	2.6	423	1 US-08-470-179-41	Sequence 41, Appli
33	30.2	2.6	2291	2 US-08-920-812-9	Sequence 9, Appli
34	30.2	2.6	2291	2 US-08-920-827-9	Sequence 9, Appli
35	30.2	2.6	2291	2 US-08-921-177-9	Sequence 9, Appli
36	30.2	2.6	2291	2 US-08-362-577C-9	Sequence 9, Appli
37	30.2	2.6	2291	3 US-08-920-828-9	Sequence 9, Appli
38	30	2.6	1478	5 US-08-817-926-1	Sequence 1, Appli
39	30	2.6	3562	5 US-08-817-926-19	Sequence 19, Appli
40	30	2.6	7742	3 US-08-882-704A-4	Sequence 4, Appli
41	29.8	2.6	970	2 US-08-690-095-2	Sequence 2, Appli
42	29.8	2.6	2275	4 US-08-743-637B-2	Sequence 2, Appli
43	29.8	2.6	2275	5 US-08-526-840B-2	Sequence 2, Appli
44	29.8	2.6	2981	2 US-08-257-073-2	Sequence 2, Appli
45	29.8	2.6	2981	3 US-08-184-009-119	Sequence 119, App

ALIGNMENTS

RESULT 1
US-07-772-087-1
; Sequence 1, Application US/07772087
; Patent No. 5275945
; GENERAL INFORMATION:
; APPLICANT: HSIAO, Hung-Yu
; APPLICANT: FODGE, Douglas W.
; APPLICANT: LALONDE, James J.
; TITLE OF INVENTION: ALKALINE PROTEASES STABLE IN HEAVY-DUTY
; TITLE OF INVENTION: DEFERENT LIQUIDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07772,087
; FILING DATE: 19911008
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16754/115 CHCO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1867 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; STRAIN: 164A-1
; FEATURE:
; NAME/KEY: CDS, 1849
; LOCATION: 716..1849
; OTHER INFORMATION: /product= "mature protein of
; OTHER INFORMATION: 164A-1"


```

REFERENCE/DOCKET NUMBER:: 1447-106P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2338 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Nephilia clavipes
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2154
OTHER INFORMATION: /product="Nephilia clavipes"
OTHER INFORMATION: dragline silk protein"
PUBLICATION INFORMATION:
AUTHORS: Xu, Ming
TITLES: Lewis, Randolph V.
TITLE: Structure of a protein superfiber: Spider
TITLE: dragline silk
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
VOLUME: 87
PAGES: 7120-7124
DATE: Sept. -1990
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 2338

```

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/781,034
FILING DATE: 19911018
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MCH91-02A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1189 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-07-781-034-4

Query Match	2.7%	Score 31.4	DB 1	length 1189
Best Local Similarity	51.0%	Pred. No. 6	8	
Matches 74	Conservative 0	Mismatches 71	Indels 0	Gaps 0

63	tgacaaatggtctgcgaagtgagccagaatctgaattgagcccaacttcataacaa	122
516	tgacgaatgtggaagaatgttcggaatccgctgtaattgacgcgaatgaagatgttcgtg	575

[illegible]

RESULT 10
PCT-US92-08328-4
Sequence 4, Application PC/TUS9208328
GENERAL INFORMATION:
APPLICANT: Fishman, Jay A.
TITLE OF INVENTION: Molecular Cloning of Antigens Shared By
TITLE OF INVENTION: Rat- and Human-Derived Pneumocystis Carinii
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millita Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/08328
FILING DATE: 19920930
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/781,034
FILING DATE: 18-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,166
FILING DATE: 30-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia

TELEFAX: (415) 857-1125
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3414 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-764-100-21

Query Match 2.7%; Score 31.4; DB 2; Length 3414;
Best Local Similarity 50.3%; Pred. No. 12;
Matches 77; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 529 tatcataaaggacatcgatcgtacgaaccggtgatacttcctgcgcggttgcaltcc 588
DB 687 TATTTCACATTCACCTGAGATTTTCACCTCTGAGAGACACTGCGCATTTAGTGTCT 628
QY 589 tgccttgcgatgtaagtaacggtgacacgtggaattcctaattcgtatgatagtcgt 648
DB 627 AGATTGTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 658
QY 649 ctgataaataattgccaataattggaatat 681
DB 567 TCCATATGATATTTCCCTTTGTTATGATTTT 535

RESULT 13

US-08-764-100-14

; Sequence 14, Application US/08764100

; Patent No. 5773700

; GENERAL INFORMATION:

; APPLICANT: van Grinsven J., Martinus Q.

; APPLICANT: De Haan, Petrus T.

; APPLICANT: Gielens L., Johannes J.

; APPLICANT: Peters, Dirk

; APPLICANT: Goldbach, Robert W.

; TITLE OF INVENTION: Improvements in or Relating to Organic

; TITLE OF INVENTION: Compounds

; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sandoz Agro, Inc

; STREET: 975 California Avenue

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/764,100

; FILING DATE: 06-DEC-1996

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/214,064

; FILING DATE:

; APPLICATION NUMBER: US 08/032,235

; FILING DATE: 17-MAR-1993

; APPLICATION NUMBER: GB 9206016.9

; FILING DATE: 19-MAR-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: No. 57737001s, Allen E.

; REGISTRATION NUMBER: 34,490

; REFERENCE/DOCKET NUMBER: 137-1061

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 354-3592

; TELEFAX: (415) 857-1125

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4970 base pairs

TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-764-100-14

Query Match 2.7%; Score 31.4; DB 2; Length 4970;
Best Local Similarity 50.3%; Pred. No. 14;
Matches 77; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 529 tatcataaaggacatcgatcgtacgaaccggtgatacttcctgcgcggttgcaltcc 588
DB 4198 TATTTCACATTCACCTGAGATTTTCACCTCTGAGAGACACTGCGCATTTAGTGTCT 4257
QY 589 tgccttgcgatgtaagtaacggtgacacgtggaattcctaattcgtatgatagtcgt 648
DB 4258 AGATTGTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4317
QY 649 ctgataaataattgccaataattggaatat 681
DB 4318 TCCATATGATATTTCCCTTTGTTATGATTTT 4350

RESULT 14

US-08-764-100-20/C

; Sequence 20, Application US/08764100

; Patent No. 5773700

; GENERAL INFORMATION:

; APPLICANT: van Grinsven J., Martinus Q.

; APPLICANT: De Haan, Petrus T.

; APPLICANT: Gielens L., Johannes J.

; APPLICANT: Peters, Dirk

; APPLICANT: Goldbach, Robert W.

; TITLE OF INVENTION: Improvements in or Relating to Organic

; TITLE OF INVENTION: Compounds

; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sandoz Agro, Inc

; STREET: 975 California Avenue

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/764,100

; FILING DATE: 06-DEC-1996

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/214,064

; FILING DATE:

; APPLICATION NUMBER: US 08/032,235

; FILING DATE: 17-MAR-1993

; APPLICATION NUMBER: GB 9206016.9

; FILING DATE: 19-MAR-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: No. 57737001s, Allen E.

; REGISTRATION NUMBER: 34,490

; REFERENCE/DOCKET NUMBER: 137-1061

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 354-3592

; TELEFAX: (415) 857-1125

; INFORMATION FOR SEQ ID NO: 20:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4970 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-764-100-20

```

Query Match 2.7%; Score 31.4; DB 2; Length 4970;
Best Local Similarity 50.3%; Pred. No. 14;
Matches 77; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

OY 529 tatcataaatgacatgcgattctgaaacccgttgatacttcgcggtgttcattcc.588
Db 773 TATTTTACAAATCCCTGGAGAAATTTCCACTCTGGAGAGAGACTGTGCCATTGTTGTCCT 714
OY 589 tgccttgcgattgagatgagaaacggatgatactctgsgaaattctaataatgcgatgtgct 648
Db 713 AGAATTGTACATGATTGCTATTAATTGTAATTTCCAACTTTCATTAATTATATAGAAATTTAGT 654
OY 649 ctgtataaatattgtcataataataattggaatat 681
Db 653 TCCATATAGATAATTTCCCTTTTCTGTATTCGATTTT 621

RESULT 15
US-08-989-478-1/c
; Sequence 1, Application US/08989478
; Patent No. 5986082
GENERAL INFORMATION:
APPLICANT: Uknes, Scott
APPLICANT: Hunt, Michelle
APPLICANT: Steiner, Henry-York
APPLICANT: Ryals, John
TITLE OF INVENTION: ALTERED FORMS OF THE NIM1 GENE CONFERRING
TITLE OF INVENTION: DISEASE RESISTANCE IN PLANTS
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5986082artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 5986082th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/989,478
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/033,177
FILING DATE: 13-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/034,379
FILING DATE: 27-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/034,382
FILING DATE: 27-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/034,730
FILING DATE: 10-JAN-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/035,021
FILING DATE: 10-JAN-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/035,022
FILING DATE: 10-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-21214/P1/GC01911
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8587
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 16, 2000, 03:09:09 ; Search time 3225.5 Seconds
(without alignments)
1359.028 Million cell updates/sec

Title: US-09-323-427-3
Perfect score: 1161
Sequence: 1 atgatgattcgtcttatgtc.....ttcgtccaatcagaagca 1161

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*

45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: gb_est33:*
53: gb_est34:*
54: gb_est35:*
55: gb_est36:*
56: gb_est37:*
57: gb_est38:*
58: gb_est39:*
59: gb_est40:*
60: gb_est41:*
61: gb_est42:*
62: gb_est43:*
63: gb_est44:*
64: gb_est45:*
65: gb_est46:*
66: gb_est47:*
67: gb_est48:*
68: gb_est49:*
69: gb_est50:*
70: gb_est51:*
71: gb_est52:*
72: gb_est53:*
73: gb_est54:*
74: gb_est55:*
75: gb_est56:*
76: gb_est57:*
77: gb_est58:*
78: gb_est59:*
79: gb_est60:*
80: gb_est61:*
81: gb_est62:*
82: gb_est63:*
83: gb_est64:*
84: gb_est65:*
85: gb_est66:*
86: gb_est67:*
87: gb_est68:*
88: gb_est69:*
89: gb_est70:*
90: gb_est71:*
91: gb_est72:*
92: gb_est73:*
93: gb_est74:*
94: gb_est75:*
95: gb_est76:*
96: gb_est77:*
97: gb_est78:*
98: gb_est79:*
99: gb_est80:*
100: gb_est81:*
101: gb_est82:*
102: gb_est83:*
103: gb_est84:*
104: gb_est85:*
105: gb_est86:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	390.2	33.6	665	42	AT11196 SMOV3MCA1
2	372.2	32.1	628	37	AA701731 SMOV3MCA1

3	364.2	31.4	672	36	AA618895	AA618895 SMOV3MCA1
4	358.6	30.9	639	44	AI322117	SMOV3MCA1
5	355.8	30.6	610	37	AA668071	SMOV3MCA1
6	324.2	27.9	776	35	AA585626	SM3D9CA34
7	312.4	26.9	537	37	AA668051	SMOV3MCA1
8	306.8	26.4	552	44	AI322068	SMOV3MCA1
9	299.6	25.8	615	44	AI317885	SMOV3MCA1
10	291.2	25.1	531	39	AA901444	SMOV3MCA1
11	278.8	24.0	363	36	AA625020	SMOV3MCA1
12	269.2	23.2	354	39	AA841200	MB3D6AA4G
13	267.8	23.1	466	36	AA625024	SMOV3MCA1
14	263	22.7	459	36	AA618952	SMOV3MCA1
15	260.4	22.4	395	36	AA625010	SMOV3MCA1
16	257	22.1	369	36	AA625022	SMOV3MCA1
17	246.8	21.3	528	44	AI322078	SMOV3MCA1
18	242.2	20.8	375	36	C71078	Yuj1
19	237.2	20.4	379	62	AV203892	AV203892
20	235.8	20.3	360	36	C67783	Yuj1
21	232.4	20.0	377	36	C70345	Yuj1
22	222.2	19.1	493	74	AM208357	SMOV3MCA1
23	215	18.5	360	62	AV201949	SMOV3MCA1
24	214	18.4	504	36	AA624955	SMOV3MCA1
25	210.8	18.2	360	62	AV196275	SMOV3MCA1
26	210.2	18.1	377	62	AV186785	SMOV3MCA1
27	207.2	17.8	360	62	AV202546	SMOV3MCA1
28	205.6	17.7	322	36	AA625021	SMOV3MCA1
29	205.4	17.7	360	62	AV201206	SMOV3MCA1
30	204.8	17.6	360	62	AV191120	SMOV3MCA1
31	184.2	15.9	417	36	AA625044	SMOV3MCA1
32	173.2	14.9	443	36	AA625040	SMOV3MCA1
33	163.2	14.1	281	46	AI438781	SMOV3MCA1
34	162.2	14.0	318	20	T01970	WEST02691 E
35	159	13.7	380	44	AI322108	SMOV3MCA1
36	146.2	12.6	635	44	AI317939	SMOV3MCA1
37	144	12.4	522	44	AI322073	SMOV3MCA1
38	140.2	12.1	317	45	AI352993	SMOV3MCA1
39	138.2	11.9	467	31	AA294218	SMOV3MCA1
40	119	10.2	273	49	AI665774	SMOV3MCA1
41	103.4	8.9	473	36	AA625005	SMOV3MCA1
42	102.6	8.8	499	44	AI313768	SMOV3MCA1
43	95.4	8.2	470	44	AI313804	SMOV3MCA1
44	87.6	7.5	409	36	AA625029	SMOV3MCA1
45	81	7.0	407	36	AA625007	SMOV3MCA1

ALIGNMENTS

RESULT 1
LOCUS AI111196 665 bp mRNA EST 31-AUG-1998
DEFINITION SMOV3MCA132SK Onchocerca volutus molting L3 larva cDNA
(SL96MLM-Ovml3) Onchocerca volutus cDNA clone SMOV3MCA1232 5',
mRNA sequence.
ACCESSION AI111196
VERSION AI111196.1 GI:3510080
KEYWORDS EST.
SOURCE Onchocerca volutus.
ORGANISM Onchocerca volutus.
Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Filarioidea; Onchocercidae; Onchocerca.

REFERENCE 1 (bases 1 to 665)
AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
TITLE Genes expressed in molting L3 larvae of Onchocerca volutus
JOURNML Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1407460.
Contact: Steven A. Williams

Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786

FEATURES
source Email: genome@smith.edu
Seq primer: Bluescript SK.
Location/Qualifiers
1..665
/organism="Onchocerca volutus"
/strain="Kumba, Camerouns"
/db_xref="taxon:6282"
/clone="SMOV3MCA1232"
/clone_lib="Onchocerca volutus molting L3 larva cDNA (SL96MLM-Ovml3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRP"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (forest strain). The L3 were cultured in 20% FCS in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volutus molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda Uni-Zap XR vector and has 1 x 10E6 independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustigman@bc.org)."

BASE COUNT 186 a 129 c 136 g 213 t 1 others
ORIGIN

Query Match 33.6%; Score 390.2; DB 42; Length 665;
Best local similarity 84.1%; Pred. No. 1.3e-94;
Matches 451; Conservative 0; Mismatches 84; Indels 1; Gaps 1;

QY 50 attcattccggttgacaatggtgacgaagtgacgaatgaaatgaaatgacgaactt 109
DB 117 ATGCTATTCCGGTGTGATTAACGCTGTAGAACGACCAAGAAATGAATGTGTCACAT 176
QY 110 caatcacatcaatttatatacagtaatgacatctgaaggaaatgtttatgaaagtc 169
DB 177 CAATTACTGTCTCACTTATATCTCGTAATCCATTGGAAGGACATATATACGTAAGGCT 236
QY 170 ttatgatacaagaaggttcgcttaatgataagtgagcgtcaagttgcggaattcac 229
DB 237 TATACGATCAGCAGATGCCAATGATGAAGTGGACGTCAGTACCGGAATTGAC 296
QY 230 ttccatttgatcaltgacaatgttgcgctacacgaatctctgaaatccagtggtattttg 289
DB 297 TTCGTTGATTCCTGTATGATGACGACGACGTCGTTGAATCCAGTGTATTTTG 356
QY 290 taacaacaacgtgttcaatcttgcattcatcattttgttacaacaatgatacgttgatc 349
DB 357 TCACATCAGTTGTTCATCTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 416
QY 350 atcgagatacaatgctttacatgtaagctgataaacaagttgagtcacagaattgagatc 409
DB 417 ATCGAATACATGCTTTTACATGGAAGCTGATNAGACAGTTAGCCCTCAACTGGAAGTT 476
QY 410 ctgaatacacaatgcttttcaaatcaaatctgtccgatacgatgacggttgcgttgcgttgcgt 469
DB 477 CCGAATACACAATTCATTCGATTCGACACACAATTCGATTCGATTCGATTCGATTCGATTCGAT 536
QY 470 ttgtgagtggtgacacacggttcaacacagttcaattgttcataatgttgtaagcagttc 529
DB 537 TTCTTGATGATGAGACATTCGACACACTGTTCATTTGTCATTCATTCATTCATTCATTCATTCAT 596
QY 530 atcataaattgatacagatcttgaacacggttgataacttctgcgcggttgatc 585
DB 597 -CCATAAATGACATGATGATTCGAAACAGTTGATACATTCCTGGGCTGTGGACAT 651

FEATURES	source
AA701731	628 bp mRNA EST 19-DEC-1997
LOCUS	
DEFINITION	SMOV3MCA1144SK Onchocerca volvulus molting L3 larva cDNA (SL96MLM-Ovm3) Onchocerca volvulus cDNA SMOV3MCA1144 5', mRNA sequence.
ACCESSION	AA701731
VERSION	AA701731.1 GI:2704931
KEYWORDS	EST.
SOURCE	Onchocerca volvulus.
ORGANISM	Onchocerca volvulus
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spirurida; Filarioidea; Onchocercidae; Onchocerca.
AUTHORS	1 (bases 1 to 628)
TITLE	Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
JOURNAL	Genes expressed in molting L3 larvae of Onchocerca volvulus
COMMENT	Unpublished (1997) On Sep 19, 1997 this sequence version replaced gi:1517220. Contact: Steven A. Williams Molecular Parasitology Smith College Department of Biological Sciences Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA Tel: 4135853826 Fax: 4135853786 Email: genome@smith.edu Seq primer: bluescript SK. Location/Qualifiers 1..628
FEATURES	
source	<pre> /organism="Onchocerca volvulus" /strain="Kumba, Cameroons" /db_xref="taxon:6282" -/clone="SMOV3MCA1144" -/clone_lib="Onchocerca volvulus molting L3 larva cDNA (SL96MLM-Ovm3)" /dev_stage="molting L3" /lab_host="X11-Blue MRF" /note="vector: lambda uni-zap XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (forest strain). The L3 were cultured in 20% FCS in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volvulus molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (mL3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda uni-zap XR vector and has 1 x 10E6 independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustig@nyc.ors)." 181 a 121 c 125 g 201 t </pre>
BASE COUNT	
ORIGIN	
Query Match	32.1%; Score 372.2; DB 37; Length 628;
Best Local Similarity	84.3%; Pred. No. 9,3e-90;
Matches 431; Conservative	0; Mismatches 78; Indels 2; Gaps 1.
0Y	50 atcgattccggttgacaaatggtgtgcgcgaagtgagcccaaaatgaatgtgaccaaactt 109
Db	112 AGGCAATTCGCGGTGATTAACGGGTAGGAAGGACCAAAATTTGATGTGCTCCAAACT 171
0Y	110 caataacaatcaatttaacacgaatgaatgcgaagacatgttaatggaagatgc 169
Db	172 CAATTACTGTCAACTTATTAATCTCGAATCCATTTTGAAGACATGATATACGCGAAAGCGT 231
0Y	170 ttatgatcaagaaggtgtcgtaaatgatgaaggtgagacgtcaagttgcggaatttcac 229

Db	232	TATAGATCAGGACAGGATGCCGAATGATGAAAGCTGACAGCTCAGGTACCGGAATGTGAC	291
Oy	230	ttccattgatcatcgtcaatgtgtcgcgblacacgaltctcgaatccacgylattttg	289
Db	232	TTCCGTTGATTCCTGTATGTAGACAGCTACACGTTCCGTTAAATACAGCTGTATTTTGG	351
Oy	290	taacacacacgtgttgatcttgcttcatlcatltaatttgltaccaaagtatgtgtcat	349
Db	332	TCACATCAGGTTGTTTCGATCTTCATTCATCCATGTTGTTGTGACAAAGTTATGCGACAT	411
Oy	350	atcgcgatcaaatgcttllacatlgagaagctcgataaacaagtlagtgcacaggtatagat	409
Db	412	ATCGAATACAAATGCTTTTTCATGAGAGCTGATTAACACGTTTACGCTTAACCTTGAAGTTT	471
Oy	410	ctgaaatcacacactgcttttccaactccaatgtgtccgagtcgcagatgcggtatgaa	469
Db	472	CCGAATATGACAACTGATTCATTTGCAACACAAATGTATCCGATGCTGTATGCGAATGTGAA	531
Oy	470	ttttgtagtggagccaacgcggtacacagttcaattgtcatcatcttggtlcaagcc--agt	527
Db	532	TTCTTGATGTGTGGACCATCCGGACCAACCTGTTTCATTTGCTACCAATGTGTCAACCCACTG	591
Oy	528	ttatcataatgacatcgcattcgcgaacc	558
Db	592	TTTACCATTAATGTGACATGTGATTGTGAACCC	622
RESULT	3		
LOCUS	AA618895	672 bp	mRNA
DEFINITION	SMOVB3MCA1879SK Onchocerca volvulus molting L3 larva cDNA (SL16MLM-Ovml3)	EST	12-NOV-1997
ACCESSION	AA618895		
VERSION	AA618895.1	GI:2522771	
KEYWORDS	EST.		
SOURCE	Onchocerca volvulus.		
ORGANISM	Onchocerca volvulus		
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida; Filarioidea; Onchocercidae; Onchocerca.		
AUTHORS	1 (bases 1 to 672)		
TITLE	Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.		
JOURNAL	Genes expressed in molting L3 larvae of Onchocerca volvulus unpublished (1997)		
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1405188. Contact: Steven A. Williams Molecular Parasitology Smith College Department of Biological Sciences Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA Tel: 4135853826 Fax: 4135853786 Email: genome@smith.edu Seq primer: plusescript SK. Location/Qualifiers 1..672		
FEATURES			
source			
	/organism="Onchocerca volvulus"		
	/strain="Kumba, Cameroons"		
	/db_xref="taxon:6282"		
	/clone="SMML3CO1879"		
	/clone="SMML3CO1879"		
	(SL16MLM-Ovml3)"		
	/dev_stage="molting L3"		
	/lab_host="XLI-Blue MR"		
	/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (forest strain). The L3 were cultured in 20% FCS in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volvulus molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (ml3). 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using		

RESULT 5
 AA668071
 LOCUS
 DEFINITION
 AA668071 610 bp mRNA EST 20-NOV-1997
 (S196MLM-Ovml3) Onchocerca volvulus molting L3 larva cDNA
 mRNA sequence.
 ACCESSION
 VERSION
 AA668071 GI:2629570
 KEYWORDS
 SOURCE
 ORGANISM
 Onchocerca volvulus.
 Onchocerca volvulus
 Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spiruridae;
 Filarioidea; Onchocercidae; Onchocerca.
 1 (bases 1 to 610)
 Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
 Genes expressed in molting L3 larvae of Onchocerca volvulus
 Unpublished (1997)
 On Oct 30, 1997 this sequence version replaced gi:2160807.
 CONTACT: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genome@smith.edu
 Seq primer: pbluescript SK.
 Location/Qualifiers
 1..610
 /organism="Onchocerca volvulus"
 /strain="Kumba, Cameroons"
 /db_xref="taxon:6282"
 /clone="SMOV3MCAM02C01"
 /clone_1id="Onchocerca volvulus molting L3 larva cDNA
 (S196MLM-Ovml3)"
 /dev_stage="molting L3"
 /lab_host="XLI-Blue MRF"
 /note="Vector: lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
 Xho I; Filarial nematode parasite of humans. Third-stage
 larvae, L3, were isolated from infected black flies in
 Cameroon (forest strain). The L3 were cultured in 20% FCS
 in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
 culture. L3 of O. volvulus molted to fourth-stage larvae by
 day 5 in culture. mRNA was isolated from approximately
 6000 molting larvae (ml3)/2000 larvae from day 1, 2 or 3
 in culture, and converted to double-stranded cDNA using
 reverse transcriptase and oligo(dt) followed by RNase H
 and DNA pol I. The library was constructed in the lambda
 Uni-Zap XR vector and has 1 x 10E6 independent
 recombinants and the average insert size is ~1200 bp. The
 library was constructed by Sara Lustigman and Michelle
 Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams.
 The library is available from Dr. Sara Lustigman (email:
 slustigmen@bc.org)."
 BASE COUNT 187 a 106 c 138 g 178 t 1 others
 ORIGIN
 Query Match 30.6%; Score 355.8; DB 37; Length 610;
 Best Local Similarity 74.0%; Pred. No. 2.4e-85;
 Matches 450; Conservative 0; Mismatches 158; Indels 0; Gaps 0;
 Oy 160 gtgaaggcttattatgacgaaggtgcgtaagtgaagtgacgtcaagtgc 219
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 2 GTAAAGGGCTTTACGATCGCATGAATGTCGTCAGATAGTGGAGCGCAGGTACT 61
 Oy 220 ggaattcaattcattatgacgtcaagtgcggtacacagatctcgaatcagct 279
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 62 GGAATGTAAGTTACATGATTCATGTAATGTGTAACATCAGGTCCTTAATCTCGA 121
 Oy 280 ggtattttgtaacaacaacgtgtgcatcttcgtttcatcattattgttacaagt 339

Db 122 GGTGCTTTGTATCAACCATTAATTGTTATTCATTTCATCCGAATTCATTACAAAAATA 181
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Oy 340 gatcgtcatatcgaatcgaatgcctttacatlgaaagctgataaacaagtgatgcag 399
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 182 GAGCAGCATATCTGATATCACTGTTTCTATATGAAAGCTGACAAACCGTTAGTCAGAA 241
 Oy 400 attgagatctgaataacacacacagtcctttcaactcaaatgtccgagtcagatgc 459
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 242 CTTGAAGTATCCGAATATACACACAGCATTCACAGACACAGTGGTACCATCTGTGCT 301
 Oy 460 cgtatgaatttggatgctgacacacacacacacacacacacacacacacacacacac 519
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 302 CGATACGAGATTTCGGAAGTGGAGACCACTGTAACCTATTTCATGCAATGATGCGA 361
 Oy 520 cagccagttatcacaataatgacacacacacacacacacacacacacacacacacac 579
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 362 GATCAGCATATATCAACAAATGAGACTTGTGATTCAGAACTACAGATATATTCGTGCAATTA 421
 Oy 580 gtccattcctcctgtcgtcgtatgataagcgtgatactgtggaattcctaagtcgat 639
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 422 GTGATTTCTTACTGTGATGACGGAAGGTGATACGATGATTCGAATGGAAGAA 481
 Oy 640 ggatgctcttgaataatctcgtcgaataatctggaataatcacaagattatgact 699
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 482 GGTGTGCTGTGACCAACTATTTACTCAACAATTTGGAATATCATCATCATATGCT 541
 Oy 700 ggccaagaagctcagctatcacaalagcgatcgatcacaagctttctatcaatgcag 759
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 542 GGCCAGAGACCTCATGTTTATTAATATACGTCGATGATGAGATTATATCATATGCAG 601
 Oy 760 atcagtat 767
 |||||
 Db 602 ATTACTAT 609
 RESULT 6
 AA585626 776 bp mRNA EST 30-DEC-1997
 LOCUS
 DEFINITION
 SM9D9CA349SK Brugia malayi L3 molting-day 9 larva cDNA
 (SM97MLM-Bml3d9) Brugia malayi cDNA clone SM9D9CA349 5', mRNA
 sequence.
 AA585626
 AA585626.1 GI:2393038
 EST.
 SOURCE
 Brugia malayi.
 ORGANISM
 Brugia malayi.
 Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spiruridae;
 Filarioidea; Onchocercidae; Brugia.
 1 (bases 1 to 776)
 Williams, S.A. and Lizotte-Waniewski, M.
 Genes expressed in day 9 post-infection, third stage larvae of
 Brugia malayi.
 Unpublished (1997)
 On May 5, 1995 this sequence version replaced gi:797827.
 CONTACT: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genome@smith.edu
 Seq primer: pbluescript SK.
 Location/Qualifiers
 1..776
 /organism="Brugia malayi"
 /db_xref="taxon:6279"
 /clone="SM9D9CA349"
 /clone_1id="Brugia malayi L3 molting-day 9 larva cDNA
 (SM97MLM-Bml3d9)"
 /dev_stage="third stage larvae, nine days after infection"
 /lab_host="E. coli XLI-Blue MRF"
 FEATURES
 source

AA625020	AA625020	363 bp	mRNA	EST	12-NOV-1997
LOCUS	SMO32MCA691SK	Onchocerca volvulus molting L3 larva cDNA			
DEFINITION	(S196MEL-00ML3) Onchocerca volvulus cDNA clone SMM13C0691 5', mRNA sequence.				
ACCESSION	AA625020				
VERSION	AA625020.1	GI:2537472			
KEYWORDS	EST.				
SOURCE	Onchocerca volvulus.				
ORGANISM	Onchocerca volvulus.				
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Filarioidea; Onchocercidae; Onchocerca.				
AUTHORS	1 (bases 1 to 363)				
TITLE	Williams.S.A., Lizotte-Maniewski.M., Laney.S. and Lustigman.S.				
JOURNAL	Genes expressed in molting L3 larvae of Onchocerca volvulus				
COMMENT	Unpublished (1997)				
	On May 9, 1995 this sequence version replaced gi:802401.				

FEATURES	Location/Qualifiers
source	1. .363

```

/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SKM13C0691"
/clone_1ib="Onchocerca volvulus molting L3 larva cDNA
(SL26MLM-Ovml3)"
/dev_stage="molting L3"
/lab_host="X11-Blue MRF"
/notes="vector: Lambda Uni-Zap XR; Site.1: Eco RI; Site.2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNA pol I. The library was constructed in the lambda
Uni-Zap XR vector and has 1 x 10E6 independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Waniewski in the laboratory of Dr. S. A. Williams
The library is available from Dr. Sara Lustigman (email:
slustig@nbc.org)."

```

Query Match	24.0%;	Score 278.8;	DB 36;	Length 363;
Best Local Similarity	85.4%;	Pred. No. 1.1e-64;		
Matches 310; Conservative	0;	Mismatches 53;	Indels 0;	Gaps 0;

[illegible]

QY	570	ctggcgaggtttccattccctctctgttcgtagatgataaggttgtaactggtggaattct	629
Db	181	ctggcctgtgtgacattcatgcttgccttggttgatgatgtaacggctgatgaaggatgttatt	240
QY	630	aatgctatgatgctgtcctctgtataaatttgcctaaataatttggataatccacaga	689
Db	241	aaatgctatgattggttgccttttggaacaattatttgcctcaacaacttggatattccacatga	300
QY	690	tttaatgctgtgccaagaagctcacgtaaacaaataatgcatgcatcaacgctttcta	749
Db	301	ttttatgctgtgacacaaagaagctcactgactacaaantatgcgcatgcatcacaactgtttcta	360
QY	750	tca	752
Db	361	cca	363

RESULT	12
AA841200	
LOCUS	
DEFINITION	354 bp mRNA EST 02-MAR-1998 M83D6AAAG1073 Brugia malayi day 6 post-infection third stage larvae SAM96MLW-BmlJd6 Brugia malayi cDNA clone 3D6AAG10 5' , mRNA sequence .
ACCESSION	AA841200
VERSION	AA841200.1 GI:2922536
KEYWORDS	EST.
SOURCE	Brugia malayi,
ORGANISM	Brugia malayi

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 354)	Blaxter,M.L., Waterfall,M., Daub,J., Licotte-Waniewski,M., Baron,L. and Jones,S.J.	Genes expressed in day six post-infection, third stage larvae of Brugia malayi	Unpublished (1997)	On Jan 15, 1998 this sequence version replaced gi:2285052.
			Contact: Blaxter ML Institute of Cell, Animal and Population Biology University of Edinburgh Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3J7, UK. Tel: +44 131 650 6760 Fax: +44 131 670 5450 Email: mark.blaxter@ed.ac.uk	
			The ABI trace of this sequence can be viewed at http://www.sanger.ac.uk/brugia/3D6/MB3D6AAG10T3.html	
			Seq primer: T3.	

FEATURES
source

```

/organism="Brugia malayi"
/strain="RRS Labs"
/db_xref="taxon:6379"
/clone="3D6AAG10"
/clone_lib="Brugia malayi day 6 post-infection third stage
larvae SAM96MLW-fml3d6"
/sex="mixed"
/dev_stage="third stage larvae, six days after infection"
/lab_host="E. coli XL1-Blue"
/note="Vector: lambdaZAPII (Unizap XR); Site 1: Eco R I
(5' end); Site 2: Xho I (3' end); Brugia malayi is a
lymphatic filarial nematode parasite of humans. mRNA was
prepared from third stage larvae of Brugia malayi isolated
from the peritoneal cavity of jirds six days after
infection. The mRNA was converted to double stranded cDNA
using reverse transcriptase and oligo(dT) followed by
RNase H and DNAPol I. The library had 2 x 10E5 independent
recombinants and average insert size was 900 base pairs.
The library was constructed by Michelle Lizotte-Waniewski.
The library is available from The Filarial Genome Project
Resource Center; contact Dr. S.A. Williams, Clark Science
Center, Smith College, Northampton, MA 01063 USA; phone +1
413 585 3826 fax +1 413 585 3786 email genome@smith.edu."

```

```

/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMML3C0758"
/clone_lib="Onchocerca volvulus molting L3 larva cDNA
(SL96MLM-OvML3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF-"
/notes="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in

```

Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith

Qy	804	acaatgttcagaacccaagaagatcggagctgttaaaacaggtggtgcccagcaaaacc	863
Db	313	ACAATGTGCAGAACCAACAGATTGAGCAGTTCGNTCAGCAATGATGATGCTGCAGC	372
Qy	864	tgc	866
Db	373	TGC	375

Search completed: April 16, 2000, 03:09:15
Job time: 6199 sec

OM of: US-09-323-427-5 to: A_Geneseq_36:* out_format: pfs
Date: Apr 16, 2000 4:42 AM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL-frame+ n2p.model -DBV=xlp  
-O/cgmr2.1/USPRO.spool/US09323427/runat_14042000_170513_19878/app_query.fasta.1  
-DB=A_Geneseq_36 -QFMT=fastan -SUFFIX=backtrans.rag  
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000  
-LOOPEXT=0.000 -GAPEXT=4.500 -GAPEXT=0.050 -GAPEXT=10.000  
-GAPEXT=0.500 -GAPEXT=6.000 -GAPEXT=7.000 -GAPEXT=10.000  
-GAPEXT=0.500 -DELOEXT=6.000 -DELEXT=7.000 -START=1  
-MATRIX=bloms62 -TRANS=human40.cdi -LIST=45 -POCALIGN=200  
-THR_SCORE=pct ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NOR-ext  
-MINLEN=0 -MAXLEN=1000000 -USER=US09323427 -NCPU=6 -ICPU=3  
-NO_XLPHY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-323-427-5  
Query length: 1161  
Database: A_Geneseq_36:*  
Database sequences: 188963  
Database length: 23686106  
Search time (sec): 139.200000
```

score_list:

Sequence	Strd	Orig	Zscore	EScore	len	Documentation
A_Geneseq_36:R20112	-	100.00	156.47	0.0329	935	AE-III (peptidylhydroxyglycine
A_Geneseq_36:P94856	-	94.50	146.24	0.1132	693	Expression plasmid pUCpIC1799 B
A_Geneseq_36:P94854	-	94.50	146.52	0.1258	875	C-terminal prepro-C-terminal al
A_Geneseq_36:R13053	-	94.50	146.52	0.1258	875	Peptidyl C-terminal alpha-amida
A_Geneseq_36:R88459	-	90.50	132.74	0.4402	1464	Feline infectious peritonitis
A_Geneseq_36:W8310	-	89.50	146.67	0.2889	374	E. coli coliclose or glucose tra
A_Geneseq_36:W13825	-	88.50	130.91	0.6650	1226	Yeast transcription regulatory
A_Geneseq_36:R06400	-	87.00	133.03	0.7727	776	Lambda gt10ch2 encoded C-termi
A_Geneseq_36:R06379	-	87.00	133.03	0.8155	866	Lambda gt10ch2 encoded C-termi
A_Geneseq_36:R06399	-	87.00	130.66	0.8640	974	Lambda gt10ch201 encoded C-term
A_Geneseq_36:R14171	-	87.00	117.43	1.51	3038	Aspergillus terreus trlri poly
A_Geneseq_36:R06477	-	86.50	130.05	0.9640	944	Protein encoded by sequence whi
A_Geneseq_36:R10323	-	85.50	130.05	0.9640	944	Type B alpha-amidating enzyme.
A_Geneseq_36:W13009	-	85.50	134.18	0.9575	560	Segment of desmosomal cadherin.
A_Geneseq_36:R20026	-	85.00	126.78	1.42	973	C-terminal amidation enzyme. cd
A_Geneseq_36:R85016	-	85.00	114.86	2.36	2710	C. difficile toxin A. Fusion P
A_Geneseq_36:W68387	-	85.00	114.86	2.36	2710	Clostridium difficile toxin A.
A_Geneseq_36:P93323	-	84.00	124.75	1.83	980	Bovine peptidyl-glycine alpha-a
A_Geneseq_36:W50144	-	83.50	122.13	2.23	1129	Oat phytochrome A apoprotein.
A_Geneseq_36:R06426	-	83.00	124.41	2.20	853	Protein encoded by sequence whi
A_Geneseq_36:R11110	-	83.00	124.41	2.20	853	Rat C-terminal amidating enzyme
A_Geneseq_36:R10322	-	83.00	124.41	2.20	853	Type A alpha-amidating enzyme.
A_Geneseq_36:R80530	-	82.00	119.95	2.66	1252	B. sphaerolus SLP. Host cell e
A_Geneseq_36:R43342	-	82.00	119.85	3.15	1068	Human p110. Recombinant polyep
A_Geneseq_36:R46294	-	82.00	119.85	3.15	1068	Ptdins 3-kinase 110 kd cataly
A_Geneseq_36:R17256	-	81.50	112.11	4.76	1911	Human PTP-OB. Protein tyrosine
A_Geneseq_36:W27235	-	81.50	112.11	4.76	1911	Human protein tyrosine phosph
A_Geneseq_36:W94027	-	81.50	112.11	4.76	1911	Human protein tyrosine phosph
A_Geneseq_36:W51405	-	81.00	130.26	4.40	369	Mouse protein-activated recept
A_Geneseq_36:R43342	-	81.00	117.90	4.05	1068	p110. Recombinant polyep
A_Geneseq_36:W63843	-	80.50	127.51	2.93	430	S. aureus asparaginyl tRNA synt
A_Geneseq_36:W1894	-	80.50	127.45	2.94	432	Asparaginyl-tRNA synthetase frc
A_Geneseq_36:W63842	-	80.50	127.45	2.94	432	S. aureus asparaginyl tRNA synt
A_Geneseq_36:P90462	-	80.50	126.72	3.03	460	Alpha-facto profilinase from S
A_Geneseq_36:W20918	-	80.50	121.77	3.44	593	H. pylori translocator protein,
A_Geneseq_36:W76041	-	80.00	111.79	5.70	1661	Hydra head activator binding p
A_Geneseq_36:R07523	-	80.00	131.15	2.73	289	Alpha-Trichosanthin encoded by
A_Geneseq_36:R29276	-	80.00	131.15	2.73	289	Ribosome inactivating protein d
A_Geneseq_36:W03474	+	80.00	127.52	3.19	385	Mouse SKI-related protein. Bovl
A_Geneseq_36:P70143	-	80.00	122.95	3.87	585	BARI gene product. New DNA frag
A_Geneseq_36:R55201	-	79.00	121.31	4.90	570	Feline zona pellucida zpb prot

A_Geneseq_36:W81811	-	79.00	121.31	4.90	570	Feline zpb protein. Isolated
A_Geneseq_36:R99462	-	79.00	102.98	10.64	2756	Biosynthetic enzyme of jcos
A_Geneseq_36:R97208	-	79.00	96.88	13.76	4655	Human calcium sensor protei
A_Geneseq_36:R97210	-	79.00	96.88	13.76	4655	Human kidney calcium sensor

seq_name: A_Geneseq_36:R20112

seq_documentation_block:

```
ID R20112 standard; Protein; 935 AA.  
AC R20112: (first entry)  
DT 06-APR-1992  
DE AE-III (peptidylhydroxyglycine N-C lyase precursor).  
KW Amidation; PHL.  
OS Xenopus laevis.  
FH Key  
FT protein location/Qualifiers  
FT 1..935 /label= AE-III  
FT /note= "including PAM and PHL domains"  
FT region 383..935  
FT /*tag= b  
FT /product= PHL  
PN EP-465404-A.  
PD 08-JAN-1992. 810399.  
PE 27-MAY-1991; JP-141678.  
PR 01-JUN-1990; JP-210535.  
PR 10-AUG-1990; JP-210535.  
PR 30-NOV-1990; JP-329911.  
PA (CIBA ) CIBA GEIGY AG.  
PI Iwasaki Y, Shimoi H, Suzuki K, Ghisalba D, Nishikawa Y;  
PI Kawahara T, Kangawa K;  
PI WPI: 92-010570/02.  
DR N-PSDB; Q20269.  
PT Novel DNA encoding peptidyl hydroxyglycine N-C lyase (PHL) -  
PT used to prepare PHL which can be used in the amidation of  
PT peptide(s) e.g. human calcitonin.  
PS Claim 4; Page 18; 28pp: English.  
CC The sequence was deduced from a cDNA insert from pAE-III-202-4  
CC (FERM BP-3172). The vector serves as a source for a DNA fragment  
CC encoding PHL for the construction of an expression vector for the  
CC prep. of recombinant PHL. The protein may be truncated to  
CC comprise only residues 383-706 or 383-713. The PHL catalyses the  
CC reaction: R-GlyOH -> R-NH2. It can be used to produce peptides with  
CC amidated C-terminal, e.g. calcitonin, growth hormone, LH-RH.  
SQ Sequence 935 AA;
```

alignment_scores:

Quality:	100.00	Length:	391
Ratio:	0.541	Gaps:	18
Percent Similarity:	47.315	Percent Identity:	19.437

alignment_block:

US-09-323-427-5/rev x R20112 ..

Align seg 1/1 to: R20112 from: 1 to: 935

1065	TGTGACCACTTCATTAACATCAATTTTAATACACGTAATGATTGCA	1016
526	CysGlnProThrAspValAlaValAspProIleThrGlyAsnhep	542
1015	AGGACATGTTTATGTAAGAGCTTTATGATCAAGAAGTGGCGTAATG	966
542	LAIAASPGLYTyr.....	546
965	ATGAAAGTGACGCTCAAGTTGCCGAATTTCACCTTCATTGATTATGC	916
547Cys 547	
915	AATGTGGCGGTACAGATCTGAATCCAGCGNATTTTGTGA.....	871
548	Asn.....SerArgIleMetGlnPheSerProAsnGlyMetHeIleMetG	563
871	871

```

563 nTpglygluGluThrSerSerAsnValProArgProGlyGlnPheArgI 580
870 .....ACAACAAGTGTGTCATTTCGTTTCATCATCATTTGTTACC 829
580 leProHisSerLeuThrMetValProAspGlnGlnLeuGlnCysValAla 596
828 AAGTTGATGTCGATATCGATACATGCTTTTACATGAGAGCTGATAA 779
597 AspArgGlnAsnGly...ArgIleGlnCysPheHisAlaGlnThrGlyAs 612
778 AACAGTTAGTCACACAGATGAGTATCTGAAATC...ACAACGCTTTTC 732
612 nPheValLys...GlnIleLysHisGlnGlnPheGlyArgGlnValPhea 628
731 AAACGTCAAAATGTCCGATGCGCATGATGCGGTTATGAAATTTTGATG 682
628 lavalSerTyAlaPro.....Gly 634
681 GGAACAACCGGTCACACGCTTCAATTTGCTATGATGTCAGCCAGTTTA 632
635 Gly.....ValLeuTyAlaValAlaAsnGlyLysProGlyTy 646
631 TCATTAATGACATGCGATTCCTGAACCGTTGATCTTTCGCGGCTTG 582
646 r.....GlyTyrSerAlaPro 652
581 TCATTCCTGCTTTGTCGATGATGCTAAGGTCATACCTGAGAAATCTTA 532
652 aIeGnGlyPheMetLeuAsnPheserSndLysP..... 663
531 AATGCTGATGATGTGCTCTGTAATATATTG.....CTAATAATT 488
664 .....IleLeuAspThrPheIleProAlaArgLysAsnPh 675
487 GGAATATCCACAGATTTAATGCTGGCCAGAAGCTCACGTTATCAAT 438
675 eaSPMePProHisAspIleAlaIaIaAspAspGlyThrValTyAlaG 692
437 ATGCGGATGATCAGAGCTTTCTATCATGCGCAGATCAGT..... 397
692 LysAspAlaHisAlaAsnAlaValITryPheSerProSerLysAlaGlu 708
396 ..ATTACCATTAAGAACAAT.....AGCGAATGTGTGCACC 359
709 HisArgSerValLysLysAlaGlyIleGluValGlnGluIleThrGluTh 725
358 ACAATGTTGGAACACACAGGATTCGAGCTGTTAAACAGGTGTCGCG 309
725 rGluIlePheGlnThrHisIleArgSerArgProLysThrAsnGlnSer 742
308 CACAACAACCTGCTGCGAGCTGCGCATGCTTTACTCAAGAAGAAAGCT 259
742 aIcIuLysGlnThrGlnGlnLysGlnLys.....GlnLysAsnSer 756
258 GCA.....GAACCGAGAAATATCATT...GATGTAGCAAC 227
757 lAaGlyValSerThrGlnGlnLysGlnAsnValValGlnGlnIleAsnAl 773
226 TGATATCAACACCCCTTGA.....ATTACGATGATATCAAGCTT 186
773 aGlyValProThrGlnGlnLysGlnAsnValValGlnGlnLysSerAlaG 790
185 TGGCACTGTGATTAGCTCAACCGTCACCTTCTGACATATGACAACT 136
790 lYalSerThrGlnGlnLysGlnSerValValGlnGlnSerSerAlaGly 806
135 GTAATACTTGTGCTACGTACAAATGGAATCTGCATCTCACATTTGGCT 86
807 ValSerThrGlnGlnLysGlnSerValValGlnGlnLysSerAlaGly 823
85 CTCATGTTTANGGTTTAAGCATTCATGATT...GCTGCCGTCATTA 39

```

```

823 lSerPheValLeuIleIleThrLeuLeuIleIleProIleAlaValLeuI 840
38 TTACCATTTGTTTAAATTCGT 16
840 leAlaIleAlaIlePheIleArg 847
seq_name: A_Geneseq_36:P94856

```

seq_documentation_block:

```

ID P94856 standard; protein: 693 AA.
AC P94856.
DT 27-JUN-1990 (first entry)
DE Expression plasmid pUCPIC1799 BglII gene product.
KW alpha-amidating; pAX799; alpha amide; ds.
OS Synthetic.
PN EP-299790-A.
PD 18-JAN-1988.
PF 15-JUL-1988; 306508.
PR 17-JUL-1987; JP-177184.
PR 05-DEC-1987; JP-306867.
PA (SUNR) Suntoxy Ltd.
PI Ohsuye K, Kitano K, Tanaka S, Matsuo H, Mizuno K;
DR WPT: 89-017279/03.
DR N-PSDB: N94527.
PT Recombinant C-terminal alpha amidating enzymes of Xenopus laevis -
PT and their precursors deoxyribonucleic acid encoding sequences.
PS Disclosure: 7pp: English.
CC The sequence encodes a derivative of the mature C-terminal alpha-
CC amidating enzyme from plasmid pAX799.
CC The plasmid was screened from an E.coli library using plasmid pXA457
CC Although pXA799 is similar to pXA457 at the N-terminus, it has an area of
CC hydrophobic elements suggesting a membrane function.
CC See also N93060.
SQ Sequence 693 AA:

```

alignment_scores:

```

Quality: 94.50 Length: 232
Ratio: 0.945 Gaps: 12
Percent Similarity: 43.103 Percent Identity: 21.552

```

alignment_block:

us-09-323-427-5/rev x P94856 ..

Align seg 1/1 to: P94856 from: 1 to: 693

```

1065 TGTGGACCACTTCAATACATCAATTTTAATACAGTAATGATCCGA 1016
491 CysGlnProThrAspValAlaValAspProIleThrGlyAsnPhePhe 507
1015 AAGCACTGTTTATGTGAAGGCTTTATGATCAAGAAGGTGCGGTAAG 966
507 lAlaAspGlyTy..... 511
965 ATGAAGGTGACGCTCAAGTTGCCGAAATTTCACTTCATTTGATTCATGC 916
512 .....Cys 512
915 AATGTTCCGCGTACAGATCTGGAATCCAGGTGATTTTGTGA..... 871
513 Asn...SerArgIleMetGlnPheSerProAsnGlyMetPheIleMetG 528
871 ..... 871
528 nTpglygluGluThrSerSerAsnLeuProArgProGlyGlnPheArgI 545
870 .....ACAACAAGTGTGTCATTTCGTTTCATCATCATTTGTTACC 829
545 leProHisSerLeuThrMetIleSerAspGlnGlnLysLeuGlnCysValAla 561
828 AAGTTGATGTCGATATGAGTACATGCTTTTACATGGAAGCTGATAA 779

```


DR WPI: 95-156754/21.
 DR N-PSDB: Q87970.
 PT Increasing protein prodn. from cultured animal cells - by adding
 PT a trichostatin to the medium, effective at low concn. and not
 PT injurious to host cells
 PS Disclosure: Page 10-15; 19pp; English.
 CC 3mu-15 cells (i.e. CHO cells expressing a peptidyl C-terminal alpha-
 CC amidase enzyme (AE)) were suspended in F-12 medium to which aliquots
 CC of trichostatin were added. Cells were cultured for 3 days at 37 deg
 CC and then assayed for AE. Without trichostatin, AE productivity was
 CC 708 U/ml. At 10, 50, 100 and 200 nM trichostatin, AE productivities
 CC were respectively 866, 1897, 1894 and 3359 U/ml.
 SQ Sequence 875 AA;

alignment_scores:
 Quality: 94.50 Length: 232
 Ratio: 0.945 Gaps: 12
 Percent Similarity: 43.103 Percent Identity: 21.552

alignment_block:

US-09-323-427-5/rev x R73053 ..

Align seg 1/1 to: R73053 from: 1 to: 875

```

1065 TGTGACCACTTCATTAACATCAATTTTAATACAGATGATTCGA 1016
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:
530 CysGlnProThrAspValAlaValAspProIleThrGlyAsnNhePheVa 546
1015 AGGACATGTTTATGTGAAGGCTTTATGATCATAGAGAGTGGCGTAATG 966
    :|||
546 lAlaAspGlyTyr..... 550
965 ATGAAGGTGAGCGTCAAGTTGCCGGAATTCACCTTCATTTGATTCATGC 916
551 .....Cys 551
915 AATGTTGCGCGTACAGATCTGGAATCCAGTGATTTTGTGA..... 871
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
552 Asn...SerArgIleMetGlnPheSerProAsnGlyMetPheIleMetG 567
871 ..... 871
567 nTPrgIyGlnIlnThrSerSerAsnLeuProArgProGlyGlnPheArgI 584
870 .....ACAACAAGTGTGTGATTCATTTGATTCATTCATTTGATTCAC 829
584 lPronHisSerLeuThrMetIleSerAspGlnGlyGlnLeuGlyValAla 600
828 AAAGTTGATCGTCATATGAGTACATGCTTTTACATGGAAGCTGATGA 779
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
601 AspArgGlnAsnGly...ArgIleGlnCysPheHisAlaIleThrGlyG 616
778 AACAGTTAGTGCAGATGAGTATGTAATC...ACAACGCTTTG 732
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
616 uPheValys...GlnIleLysHisGlnIlePheGlyArgGlnValAlaPhe 632
731 AAACCTCAATTTGTCGCGATGCCAGTATGCCGTTATGAATTTTGCATG 682
    :|||:|||||
632 lavalSerIyrAlaPro.....Gly 638
681 GGACCAACCGGTCAACAGATTCATTTGCTATCATTTGTCAGCCAGTTGA 632
    ||| :|||:|||||:|||||:|||||:|||||:|||||:
639 Gly.....ValLeuIyrAlaValAlaAsnGlyLysProIleTyr 650
631 TCATTAATGACATGCGATTCGAAACCGTGTACTTCTGCGCGGTG 582
    :|||:|||||
650 rGly.....AspSerThrProValGlnGlyPhe..... 659
581 TCCATTCCTGCTTGTGATGATGAACGCGTATGATCTGGAATTTCTTA 532
    :|||:|||||
660 .....MetLeuAsnPheSerAsnGlyAsp..... 667

```

531 AATGCTGATGATGCTGCTGTGATTAATATTG.....CTAATATATT 488
 668IleLeuAspThrPheIleProAlaArgLysAsn 679
 487 GGAATATCCACAGATTTAATGCTGGCCAGAGAGCTCACGATAC 442
 679 eGlnMetProHisAspIleAlaIleAlaLysAspGlyThrValTyr 694

seq_name: A_Geneseq_36:R88469

seq_documentation_block:

ID R88469 standard; Protein; 1464 AA.
 AC R88469.
 DT 14-AUG-1996 (first entry)
 DE Feline infectious peritonitis 1 virus spike protein.
 KW Feline infectious peritonitis 1 virus; FIPV-I; spike protein;
 OS Feline infectious peritonitis 1 virus.
 PN J07327683-A.
 PD 19-DEC-1995.
 PF 10-JUN-1994; 129300.
 PR 10-JUN-1994; JP-129300.
 RA (KITA.) KITASATO KENKYUSHO SH.
 DR WPI: 96-072341/08.
 DR N-PSDB: T10166.
 PT DNA encoding feline infectious peritonitis 1 virus spike protein -
 PT used in a vaccine for prevention and treatment of FIPV-I infection
 PS Claim 1: Page 14-17, 23pp; Japanese.
 CC This sequence represents the feline infectious peritonitis 1 virus
 CC (FIPV-I) spike protein. The FIPV-I spike protein may be used in the
 CC production of a vaccine for the prevention and treatment of FIPV-I
 CC infection. The spike protein may be produced by transforming a host
 CC cell with the spike protein DNA and expressing the sequence such
 CC that the spike protein can be isolated.
 SQ Sequence 1464 AA;

alignment_scores:
 Quality: 90.50 Length: 382
 Ratio: 0.508 Gaps: 16
 Percent Similarity: 46.597 Percent Identity: 19.895

alignment_block:

US-09-323-427-5/rev x R88469 ..

Align seg 1/1 to: R88469 from: 1 to: 1464

```

1029 CGTAATGCAATTCGAAGACATGTTTATGTGA.....GCTCTTTA 989
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
739 LysAsnSerThrThrGlyIleIlePheThrValAlaProCysAspLeuTh 755
988 TGATCAGAAGGTTGCCGTAATGATGAGTGCAGCTCAAGTTGCCGGA 939
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
755 lAlaGlnValAlaValAlaIleAsnAspGlnIleValGlyAlaIleThrAlaV 772
938 TTTCACTTCCA.....TTTGATTCATGCAAT.....GTTCGCGCT 904
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
772 lAlaAsnIlnThrAspLeuPheGlnPheValAlaAsnThrGlnAlaValG 788
903 ACACGATCTCTGAATCCAGCTGATTTTGTGTAACACACACTGTGTCAT 854
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
789 SerArgSerSerThrProAsn.....PheValThrSerTyrThrMetPr 803
853 TTTCGTTTCACTTATTTGTTACCAAAAGTTGATGCGATATCGAGTAC 804
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
803 oGlnPheIyr.....TyrIleThrLysTrpAsnAsnAspThrSerSerA 818
803 AATGCTTTTACATGGAAGCTGATAAACAAGTATGTCACAGATTGAGTA 754
    :|||:|||||
818 snCys.....ThrSerAlaIleThrIyr 825
753 TGTGAATACACACTGCTTTTCAAACTCAATTTGCCGATGCGCATGANG 704
    :|||:|||||:|||||

```

```

826 SerSerPheAlaIleCysAsnThrGlyIuIleGlyTyrValAsnValn 842
703 CCGTATGAAATTTGGATGCT ..... 682
842 rHisValGluIleValAspSerIleGlyValIleLysProValSerT 859
681 ..GGACCAACCGGTCAACCGATTCAATTTGATCATTTGGTCAGCCAGTT 634
859 hngIAsnIleSerIleProLysAsnPhetHrValAlaValGlnAlaGlu 875
633 TTTCAATTAATGACATGCATCTTCAACCGTTGACTTTCTGGCGGT 584
876 TTTTleGlnIleGlnValLysProValValAsp.....CysAlaTh 890
583 TGTCCATTCCTGCTTTGTCATGATGATGTAACGGTGTACTGTGAATTC 534
890 r.....TyrValCysAsnGlyAsnThrHisCysLeuLysLeu 903
533 TA.....AATGCT.....GAT 523
903 euThrGlnTyrThrSerAlaCysGlnThrIleGlnAsnAlaLeuAsnLeu 919
522 GGATGTCCTTGATTAATTTTGCATAATTTTGGAAATATCCACACAGA 473
920 GtYAlaArgLeuGlnSerLeuMetLeuAsnAspMet..... 931
472 TTTAATGGCTGCCAAGAAGTCTACGTATACAAATATGCGGATCATAC 423
931 ..... 931
422 ACCTTTTCTATCATGCCAGATCCAGTATTACATTAAGAACCAATATAC 373
932 .....IleThrValSerAspArgLysLeu 939
372 GAATGTCCTGCACACAAATGTCACAGACCAACAGATTCGAGCGCTTAA 323
940 GluLeuAlaThrValGlnArgPheAsnAlaThrAlaLeuGlyGlyIuLys 956
322 AACAGGTGGTCCGCCAGCAAAACCTGCTGACCTGCGCAACTTCCTTAC 273
956 sleuGlyLeuTyrPheAspGlyLeuSerSerLeuLeuProProLysI 973
272 TCAAGAAAAGATCTGCAGAACCGAGAAATATCATTTGAT.....GTACA 229
973 leGlyLysArgSerAlaValGlnAspLeuPheAsnLysValValThr 989
228 ACTGATATACAAACCTTGAATTAAGCATGATATACAGCTTTGGCACT 179
990 SerGlyLeuGlyThrValAspAspArgPyrLysLysCysSerSerGlyTh 1006
178 TGATTTACCTCACCGTCTGCAATCATATGATGACAAACCTGTATATAC 129
1006 rAspValAlaAspLeuValCysAlaGlnTyrTyrAsnGlyIleMetValL 1023
128 TTGCTGCAGTACAAATGGAATTCGATGCATCCATTTGGCTTCAATG 79
1023 euProGlyValAlaAspGlyAsnLysMetSerMetLysThrAlaSerLeu 1039
78 TTTATGGGTTTAAGCATT..GCATTTGATTTGCTCCGCTATATATCCAT 32
1040 lIleGlyGlyMetAlaLeuGlySerIleThrSerAlaValAlaValAla 1056
31 TT.....CGTTTAATTTTGTCCCAATCAGA 6
1056 heAlaMetGlnValGlnAlaArgLeuAsnTyrValAlaLeuGln 1070
seq_name: A_Geneseq_36:W88310
seq_documentation_block:
ID W88310 standard; Protein; 374 AA.
AC W88310:
DE 26-APR-1999 (first entry)
E. coli collitose or glucose transferase.

```

```

KW O antigen; O111 antigen; wbdm gene; collitose transferase;
KM glucose transferase; diarrhoea; haemorrhagic colitis; diagnosis.
OS Escherichia coli.
PN W09850531-A1.
PD 12-NOV-1998.
PE 01-MAY-1998; AU0315.
PR 22-JUL-1997; AU-008162.
PR 01-MAY-1997; AU-006545.
PA (UNSY ) UNIV SYDNEY.
PI Reeves PR, Wang L;
DR WPI: 99-059669/05.
DR N-PSDB; X06748.
PT Nucleic acid molecules specific for bacterial polysaccharide
PT antigens - useful for detecting specific strains in, e.g. food,
PT faeces or patient samples
PS Disclosure: Fig 7, 16pp; English.
CC This is the amino acid sequence of the protein encoded by the wbdm
CC gene of a gene cluster (see X06748) involved in the biosynthesis of
CC the Escherichia coli O111 antigen. The protein shows high
CC homology with TSE of Yersinia enterocolitica, and is predicted to
CC be a collitose or glucose transferase. The use of nucleic acid
CC molecules derived from particular assembly and transport genes,
CC particularly wbd (transferase), wzx (flippase) and wzy (polymerase)
CC genes, within O antigen gene clusters improves the specificity of
CC methods for the detection and identification of O antigens, e.g. in
CC testing food- or faecal-derived samples, or samples from patients.
CC The O antigen is a major virulence factor of enteropathogenic E.
CC coli strains that cause diarrhoea and haemorrhagic colitis.
SQ Sequence 374 AA;

```

```

alignment_scores:
  Quality: 89.50      Length: 363
  Ratio: 0.533      Gaps: 18
  Percent Similarity: 46.281      Percent Identity: 21.763

```

alignment_block:

US-09-323-427-5/rev x W88310 ..

Align seg 1/1 to: W88310 from: 1 to: 374

```

1100 TTGACAAATGCTGTGCAAGGTGACGCCAAGAAATTTGATGTGACCAACTTCA 1051
||||| :|||:|||||:|||||:|||||:|||||:
38 LeuThrGlyIleSerGluValLysProThrGlnAsnIleAsn..... 51
1050 ATAACATCAATTTTATACAGTAAATGCATTCGAAGAC.....A 1010
52 .....| ||||| :|||:
1009 TGTATATGTGAAGGCTTTTATGATCAAGAGTTCCTGTAATGATGAG 960
||||| :|||:|||||:|||||:|||||:|||||:
63 erPheheArgAlaLeuPheGlnValLysIleLeValAlaLeuLys 79
959 GTGACGTCAAGTTGCCGAATTTCACTTTCATTTGATTCATGCAATGTT 910
||||| :|||:|||||:|||||:|||||:
80 ProAspIle.....IleHisSerHisMetPheHisAlaAsn.. 91
909 GCGCGTACACGATCTGTGAATCCACGTGATTTTGTGAACAACAAGTGT 860
92 .....I 92
859 TGTCAATTTGTTTCATTCATTAATTTGTTACCAAGTTG...ATCGTGCAT 813
||||| :|||:|||||:|||||:|||||:|||||:
92 lePheSerArgPheIleArgMetLeuIleProAlaValProLeuIleCys 108
812 ATGCAGTACAAATGCTTTTATCATGGAAGCTGATAAACACTTATGTGCACAG 763
||||| :|||:|||||:|||||:|||||:
109 ThrAlaHisAsnLysAsnGluGlyGlyAsnAlaArgMetPheCysTyrIa 125
762 ATTGAGGTATTCGAATGCACACTGCTTTTCAACGCAATTTGTCCCGGAT 713
||||| :|||:|||||:|||||:|||||:
125 glu.....SerAspPheLeuAlaSerIleThrThrAsnValS 138

```

```

712 GCCAGTATGCCGTTATGAAATTTGGATGGTGGACCAACCGGTCAACAG 663
    :: |||||:::|||||:::
138 eelysgluValaValaIngluPhe.....llealaArgylsAlaThrPro 152
    |||:::
662 TTCAATTTGCTATCATTGGTCACCGCACTTTATCAT..... 628
    |||:::
153 LysAsnLyslleValaGluThrProAsnPheIleAsnThrAsnLysPheAs 169
627 .....AAATGACATCGATTCCTGGAACGGTTGATA 597
169 pheAspIleAsnValaArgLysThrArgAspAlaPheAsnLysLysA 166
596 CTTTCGCGGGGTTCCTCATCTCTGTCGATGATGTAACGGTGAT 547
    |||||:::
186 spserThrAlaValleuLeuAlaVal.....GlyArg 196
546 ACTGTGGA.....ATTCTAATGCTGATGATGCTGC 515
    |||||:::
197 LeuValGluValaLysAspTyrProAsnLeuLeuAsn.....Al 209
514 TCTTGATTAATATTGCTAATAATATTGAAATATCCACAGAT..... 472
    |||||:::
209 aileAsnHisLeuIleleuSerLysThrSerAsnCysAsnAspPheIleL 226
471 ..TTAATGCTGGCCAAGAAGCTCAGCTATACAAATATGCGGATCGATCA 424
    |||||:::
226 euleuIleAlaGlyAspGlyAlaLeuAlaArgAsnLysLeuLeuAsp..... 240
423 CAGCTTTTCATCAATGCCAGATCAGATATTACCATTA..... 385
    |||||:::
241 .....LeuValCysGlnLeuAsnLeuValAspLysValPhePheLe 254
384 .GAACCAATAGCGAATGTGTCACCAATGT..... 352
    |||||:::
254 uGlyGlnArgSerAspIleLysGlnLeuMetCysAlaAlaAspLeuPheV 271
351 .....TCAGAACAACAAGATTGCGA...GCTGTTAAACAGGTGCT 313
    |||||:::
271 alleuSerSerGluThrGluGlyPheGlyLeuValAlaIleGluAlaMet 287
312 GCCGCAGCAAAACCTGCTGACGTGGCAACTTCGTTTACTCAAGAAAG 253
    |||||:::
288 AlaCysGluArgProValAlaIleThrAspSerGlyGlyAlaLysGluVa 304
262 ATCTGAGAACCGGAGAAATATCATGTATGATGACAGATGATC..... 220
    |||||:::
304 ValaGlyProHisAsnAspValIleProValSerAsnHisIleLeuLeuA 321
219 .....AACACCTTGAATTAAGCATGATATCAAGCTTGG 184
    |||||:::
321 laGluLysIleAlaGluThrLeuLysIle...AspAspAsnAlaArgLys 326
183 CCAGTTGATTAGTCACCGTCGACTTCGCACATCAAT 145
    |||||:::
337 lleIleGlyMetLysAsnArgGluTyrIleValSerAsn 349
seq_name: A_Geneseq_36:W13825
seq_documentation_block:
ID W13825 standard; Protein: 1226 AA.
AC W13825;
DT 04-JUN-1997 (first entry)
DE Yeast transcription regulatory factor SRB8.
KW Transcription regulatory factor; suppressor of RNA polymerase B;
KW SRB8; RNA polymerase II; holoenzyme; SWI/SNF.
OS Saccharomycetes cerevisiae.
PN W09708301-A1.
PD 06-MAR-1997.
PE 28-AUG-1996; U14192.
PR 31-AUG-1995; US-521872.
PR 11-OCT-1995; US-540804.
PR 26-JAN-1996; US-590399.
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.

```

```

PI Chao DM, Koleske AJ, Thompson CM, Young RA;
DR WPI: 97-179258/16.
DR N-PSDB: T59908.
PT Purified RNA polymerase II holoenzyme - comprises RNA polymerase II
PT and one or more regulatory proteins, pref. suppressor of RNA
PT polymerase B proteins or SWI/SNF proteins
PS Claim 11: Fig 10a-b; 14pp; English.
CC Novel yeast SRB (suppressor of RNA polymerase B) proteins SRB4, SRB5,
CC SRB6, SRB7, SRB8, SRB9, SRB10 and SRB11 (W13821-28) are transcription
CC regulatory factors that act as positive and negative regulators of
CC RNA polymerase II activity, and are components of the RNA polymerase
CC II holoenzyme. They were identified using methods designed to
CC identify transcription factors involved in RNA polymerase II
CC C-terminal domain (CTD) function. SRB8 and SRB9 appear to repress
CC CTD activity. Genomic clones (T59904-11) for the SRBs have been obd.
CC SRBs can be used to treat diseases resulting from alteration or
CC deletion of the SRB gene, pref. by gene transfer technology. They
CC can also be used in vitro transcription of DNA and to identify
CC cpds. that modify gene transcription.
SQ Sequence 1226 AA;

alignment_scores:
    Quality: 88.50      Length: 391
    Ratio: 0.468      Gaps: 16
Percent Similarity: 48.338      Percent Identity: 20.716

alignment_block:
US-09-323-427-5/rev x W13825 ..

Align seg 1/1 to: W13825 from: 1 to: 1226

1148 TTAATGCTTCTGCTACTACCACTTATGCAATGCTTATTCGATTCGCGTT 1099
    ::|||:::|||||:::
358 MetMetProSerIleuTyrArgLeuLeuAsnIleLeuIleThrTyrGlyI 374
1098 GACAAATGCTGTGCAAGTGCAGCCAGAATTGAATGTGACCAACTTCAAT 1049
    :|||:::
374 eileLysValaProThrTyrIleArgLysLeuIle..... 385
1048 AACATCAATTTTAATACACGTAATGCTATTCGACAGCATGTTTATGTGA 999
    |||||:::
386 ..SerSerGlyLeuLeuTyrLeuGlnAspSerAsnAspLysPheValHis 401
998 AAGCTTTATGATCAAG...AAGCTTGCCGTAATGATGATGAGTGACGT 952
402 ValGlnLeuLeuIleAsnLeuLysIleSerProLeuMetLys..... 415
951 CAAGTTGCCGGAATTTCACTTCATTTGATTCATGCAATGTTGCCGGTAC 902
    |||||:::
416 .....SerGlnTyrAsnMetValleuArgAsnValMetGluT 428
901 AGCATCTCGAATCCACGTGATTT..... 876
    |||||:::
428 yAspValLysPheTyrGluIlePheAsnPheAspGlnValaGluIle 444
875 .....TTGTAACAACAACAGTGTGTCATTTGTTTCATCATCATTAATT 835
    |||||:::
445 ThrGluGlnIleLysMetArgIleLeuSerAsnAspIleThrAsnLeuG 461
834 GTTACCAAGATTGATGCTGATGATGAGTACATGCTTTTACATGAGAAC 785
    |||||:::
461 neuSerLysThrProLeuSerIleLysIleMetValaIleGluTyrPyl 478
784 TG.....ATMAACAGTTAGTGACACGATTTGAGAGTAT 753
    |||||:::
478 euSerHisLeuCysSerGlyIleLeuSerSerValAsnArgThrValLeu 494
752 CTGAATACACACTGCTTTTC.....AACTCAATTTGCCGATGCCA 709
    |||||:::
495 LeuLysIlePheLysIlePheCysIleAspLeuGluValaPheHisHissh 511
708 GTATGCCGTTATGAATTTTGGATGGTGGACCAACCGGTCAACAGTTCA 659

```

```

511 eheuystrpiledglupheilevaltyrhsglnleuSerAspIleg 528
518 ATTTGCTA.....TCATTGCTC 642
528 IuSerleuInuAlaleuMetAspIleLeuLeuCysTyrGlnIuSerLeuPhe 544
641 ACCGATTATCATTAATGACATGCGATTCTG...AACCGTGATA.. 597
545 SerGlnPhe...IleAsnAspHisIleLeuPheThrIlePheIlePhe 560
596 .....CTTCTGCGCGGTTCATTCCTCTGCTTGTGATGA 560
560 eiletyrIuIysValIleuIysGlnIuIysAspValProAlaIyrAsnValT 577
559 TGTACAGGTGATCTGTGGAATTCCTAATGCTGATGATGCTGCTTG 510
577 hSerPheMetProPheIyrIysPhePheMetIysAsnPheProPheVal 593
509 ATAAATATTTGCTAAATATTAATTTGAATATCCAAAGATTTAATGCTGCGC 460
594 LeuIysValAspAsnAspIleuArIlegIuLeuGlnSerValIyrAsnAs 610
459 CAAGAAGCTCAGCTATACAAATATGCGGATGATCAGCTTTTCTATCA 410
610 pEluIysLeuIysTThGlnIuIysLeuIysAsnAspIysSerGlnValLeu 627
409 ATGCCAGATCATATTACCATTAAGAACCAG..... 378
627 yValIyrSerMetIleAsnAsnSerAsnGlnAlaValAlIylGlnThrTP 643
377 ..ATAGCGAATGTGTGACCAACCATGTTTCAG.....AACCAAGCA 337
644 AsnPheProGlnValAlPheGlnValAsnIleArPheLeuLeuHis...As 659
336 TTTCGAGCTGTTAAACACAGGTGTCGCCAGCAA..... 303
659 nSerGlnIleIleAspThrAsnThrSerIysGlnPheGlnIysAlaArgA 676
302 .....AACCTGCTGACGCTGCGCAACTTC 279
676 snAsnValMetLeuLeuIleAlaThrAsnLeuIysGlnIyrAsnIysPhe 692
278 GTT.....TACTCAAGAAAGATCTGCAGAACCGAGATA..... 243
693 MetSerIlePheLeuIysArIysAspPheThrAsnIysAsnLeuIlegI 709
242 .....TCATTGATGTAGCACTGATATCAACACCCCTGA 209
709 nLeuIleSerLeuIysIleuLeuThrPheGlnValIThrGlnAsnValLeuG 726
208 AATTAGCGATGATTAATCAGCTT 186
726 IyLeuGlnIyrIleIleArGleu 733

```

seq_name: A_Geneseq_36:R06400

seq_documentation_block:

ID R06400 standard; protein; 776 AA.
AC R06400;
DT 17-DEC-1990 (first entry)
DE Lambda gt10ch2 encoded C-terminal alpha amidating enzyme.
KW C-terminal alpha amidating enzyme; human thyroid gland;
KW lambda gt10ch2.
OS Homo sapiens.
PN W09008190-A.
PD 26-JUL-1990.
PE 17-JAN-1990; J00042.
PR 17-JAN-1989; JP-005878.
PA (SUNR) SUNTORI LTD.
PI Ohnuye K, Kitano K, Tanaka S;
DR WPI: 90-254034/33.
DR N-PSDB: 005631.

PT C-terminal alpha amidating enzymes from human thyroid gland -
PT converting C-terminal glycol peptide(s) or protein, and new DNA
PT coding sequences.
PS Disclosures: pp; English.
CC cDNA libraries were prepared from human thyroid gland poly(A) RNA
CC in lambda gt10 and gt10. These were screened using probes derived
CC from DNA encoding similar enzymes in Xenopus laevis. Three clones
CC were isolated from the gt10 library, including gt10ch2.
CC Restriction analysis indicated that gt10ch2 encoded a different
CC type of enzyme to one of the other two clones. The DNA can be
CC inserted into vectors for expression in E.coli or (more efficiently)
CC in animal cells.
CC The enzyme is useful for prodn. of physiologically active alpha-
CC amidated peptides from their C-glycol precursors.
CC See also 005630 and 005632.
SQ Sequence 776 AA;

alignment_scores:

Quality:	87.00	Length:	322
Ratio:	0.592	Gaps:	13
Percent Similarity:	45.652	Percent Identity:	18.944

alignment_block:

us-09-323-427-5/rev x R06400 ..

Align seg 1/1 to: R06400 from: 1 to: 776

```

1065 TGTGACCACTTCANTAAATCAATTTTAATACAGTAATGATTCGA 1016
    ||| |||||.....: |||
434 CysGlnProThrAspValAlaValAspProGlyThr..... 445
1015 AGGACATGTTATGGAAGGCTTTATGATCAAGAAGTCCCGTAAG 966
    ||| |||||.....: |||
446 .GlyAlaIleIyrValSerAspIyTr..... 454
965 ATGAAGGTGACGCTCAAGTTGCCGGAATTCATTCCATTGATTGATGCG 916
    |||
455 .....Cys 455
915 AATGTGGCCGTACAGCATCTGAAATCCAGCTGATTTGTGAACAC 866
    ||| |||||.....: |||
456 Asn...SerArgIleValGlnPheSerProSerGlyIysPheIleThrG1 471
865 A.....ACTG 861
471 nTrpGlnIuGlnIuSerSerGlySerSerProLeuProGlnIuPheThrV 488
860 TT.....GTCAATTGCTTCATCCATTAATT.....GTACCAAAAGTT 823
    ||| |||||.....: |||
488 AlProHisSerLeuAlaLeuValProLeuLeuGlnIuLeuCysValAla 504
822 GATCGTGCAATAT...CGATACAAATGCTTTTACATGGAAGCTGATTAAC 776
    ||||| |||||.....: |||
505 AspArgGlnAsnGlnIyArgIlegIncIysPheIysThrAspThrIysGlnPhe 521
775 AGTTAGTCACAGATGAGGTATCTGAATTCACAACTGCTTTTCAACAC 726
    ||| |||||.....: |||
521 eValArgGlnIuIeIysHisSerSerPheGlyArgAsnValPheAlaIylS 538
725 AAATGTCGCCGATGCCATATGCCGTTATGAATTTGGATGGTGAGCCA 676
    ||| |||||.....: |||
538 eTyrIlePro..... 541
675 ACCGGTCAACCAAGTTCAATTTGCTATCATTTGTCAGCCAGTTTATGATA 626
    ||| |||||.....: |||
542 .....GlyLeuPhePheAlaValAsnGlnIyIysProHisPheIy.. 554
625 ATGACATCGGATTCGAAACGTTGATCTTCTGCGGGTGTGCAT 576
    ||||| |||||.....: |||
555 .....AspGlnIuProValGlnIyPhe..... 562
575 CCTGTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 526

```

```

563      ....:Valmetaspheserhsnldylutellelsplle..... 575
525 GATGATGTCTCTGATAAATATTTGCAATATTCGAATTCGAAC 476
576      ....:LysprovalaIrgLgnshApsmerprohI 586
475 AGATTATTAAGGCTGGCCAGCAAGCTCAGCTATTCGAATATTCGAGATTCAT 426
586 ssapllevalAlasergluaasprgylthValTYrlleglasprAlahNST 603
425 CACACCTTTTCATCAATGCAGATCAGTATTCACATTAAGACCAAT 376
603 hAsnThrValTrpLysphenrleuthrglLulysleuglnhIsatserg 619
375 ACCGAATGTGTTCCACACCAATGTTCCGAACCCACAGATTCGAGCTGT 326
620 ValTysLysAlaGlyllegluaIvalnglnlulLysgluaAlaIaVa 636
325 TAAACAGGTGGTGGCCGACGACCAACCTGCTGGAGCTGGCAACTTCGTT 276
636 lValgluthrLysmetgluaasnLysproInrserSergluleuglnLysM 653
275 TACTCAGAAAGAATCT.....GCAGACCGGAGAAATATCATGTATGA 232
653 eGlnglulysgluLysleuLysleuLysLuproclYserglYalProval 669
231 CGAACTGATATCACACCCCTTGAATATGGCAGATGATATCAAGCTTTGCC 182
670 .ValleuLeThrThrLeuLeuValIleProvalValValLeuLeuAlaI 686
181 AGTTGATTTCATGC 168
686 leAlallehele 690

seq_name: A_Geneseq_36:R06379

seq_documentation_block:
ID R06379 standard; protein: 866 AA.
AC R06379:
DT 17-DEC-1990 (first entry)
DE Lambda g11och101 encoded C-terminal alpha amidating enzyme.
KW C-terminal alpha amidating enzyme; human thyroid gland;
OS Homo sapiens.
PN WC9008190-A.
PD 26-Jul-1990.
PE 17-JAN-1990; J00042.
PR 17-JAN-1989; JP-005878.
PA (SONR ) SUNTORY LTD.
PI Ohshuye K, Kitano K, Tanaka S;
DR WPI: 90-254034/33.
DR N-PSDB: 005632.
PT C-terminal alpha amidating enzymes from human thyroid gland -
PT converting C-terminal glycyI peptide(s) or protein, and new DNA
PT coding sequences.
PS Disclosure; pp: English.
CC cDNA libraries were prepared from human thyroid gland poly(A) RNA
CC in lambda g11 and g110. These were screened using probes derived
CC from DNA encoding similar enzymes in Xenopus laevis. Three clones
CC were isolated from the g110 library, including g11och101.
CC Restriction analysis indicated that g11och101 encodes a different
CC type of enzyme to one of the other two clones. The DNA can be
CC inserted into vectors for expression in E.coli or (more efficiently)
CC in animal cells.
CC The enzymes are useful for prodn. of physiologically active alpha-
CC amidated peptides from their C-glycyI precursors.
CC See also Q05630 and Q05631.
SQ Sequence 866 AA;

```

```
alignment_scores:
  Quality: 87.00  Length: 322
```

```

Ratio: 0.592      Gaps: 13
Percent Similarity: 45.652      Percent Identity: 18.944

alignment_block:
US-09-323-427-5/rev x R06379  ..

```

Align seg 1/1 to: R06379 from: 1 to: 866

[illegible]


```

726 |ValGluThrIlysmetGluAsnIlysProThrSerSerGluLeuGlnIlysm 743
275 |TACTCAGAAAGATCT.....GCAGAACCGGAGAAATATCATGTATGTA 222
743 |etGlnIlysglnIlyLeuIleIlysglnProGlySerGlyValProVal 759
231 |CGAAGTATATCAACACCCCTTGAAATTAAGCATGATTAATCAAGCTTTGCC 182
760 |ValLeuIleThrThrLeuLeuValIleProValValValLeuLeuAlaI 776
181 |AGTTGATTTAAGTC 168
776 |LeaIlePheIle 780

```

seq_name: A_Geneseq_36:R06399

seq_documentation_block:

```

ID R06399 standard; protein: 974 AA.
AC R06399;
DR 17-DEC-1990 (first entry)
DE Lambda gt10ch201 encoded C-terminal alpha amidating enzyme.
KW C-terminal alpha amidating enzyme; human thyroid gland;
KM lambda gt10ch201.
OS Homo sapiens.
PN WO9008190-A.
PD 26-JUL-1990.
PF 17-JAN-1990; J00042.
PR 17-JAN-1989; JP-005878.
PA (SUNK) SUNKORY LTD.
PI Ohnuye K, Kitano K, Tanaka S;
DR MPI; 90-254034/33.
DR N-PSDB; 005630.
PT C-terminal alpha amidating enzymes from human thyroid gland -
  converting C-terminal glycol peptide(s) or protein, and new DNA
  coding sequences.
PS Disclosure; pp: English.
CC cDNA libraries were prepared from human thyroid gland poly(A) RNA
  in lambda gt11 and gt10. These were screened using probes derived
  from DNA encoding similar enzymes in Xenopus laevis. Three clones
  were isolated from the gt10 library, including gt10ch201.
CC Restriction analysis indicated that gt10ch201 encodes a different
  type of enzyme to the other two clones. The DNA can be inserted
  into vectors for expression in E.coli or (more efficiently) in
  CC animal cells.
CC This enzyme is useful for prodn. of physiologically active alpha-
  amidated peptides from their C-glycyl precursors.
CC See also Q05631 and Q05632.
SQ Sequence 974 AA:

```

alignment_scores:

```

Quality: 87.00 Length: 322
Ratio: 0.592 Gaps: 13
Percent Similarity: 45.652 Percent Identity: 18.944

```

alignment_block:

US-09-323-427-5/rev x R06399 ..

Align seg 1/1 to: R06399 from: 1 to: 974

```

1065 TGTGACCACTTCATTAACATCAATTTAATACAGTAATGCATTGCA 1016
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:
631 CysGlnProThrAspValAlaValAspProGlyThr.....642
1015 AGGACATGTTTATGAAAGCTCTTATGATCAAGAAGTTGCCGTAAATG 966
    ||| ::|||:|||||:|||||:|||||:|||||:|||||:
643 .GlyAlaIleThrValSerAspGlyTyr.....651
965 ATGAAGTGCAGCTCAAGTTGCCGGAATTTCACTTCATTTGATTATGC 916
    ||| |||:|||||:|||||:|||||:|||||:|||||:
652 .....Cys 652

```

```

915 AATGTGCGGGTACAGACCTCTGAATCCAGCGTATTTTGTAAACAC 866
    ||| ::|||:|||||:|||||:|||||:|||||:|||||:
653 Asn...SerAlaIleValGlnPheSerProSerGlyLysPheIleThrG 668
865 A.....ACTG 861
668 nPrGluGlnIlySerSerGlySerSerProLeuProGluGlnPheThrV 685
860 TT.....GTCAATTGCTTTATCCATTTAATT.....GTAAACAAAGT 823
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
685 aLProIleSerLeuAlaLeuValProLeuLeuGlyGlnLeuCysValAla 701
822 GATCGTCATAT...CGAGTACATGCTTTTACATGAGTGAAGTAAAC 776
    ||||| |||||:|||||:|||||:|||||:|||||:|||||:
702 AspArgGluAsnGlyArgIleGlnIlySerPheIlyThrAspThrIlysglu 718
775 AGTTAGTGCACAGATTGAGGTATGTGAATTCACAACTGCTTTCAACATG 726
    ||| ::|||:|||||:|||||:|||||:|||||:|||||:
718 eValArgGluIleLysIleSerSerPheGlyArgAsnValPheAlaIleS 735
725 AATTCGTCGCGATGCCAGTATGCCGTTATGAATTTTGGATGCGTGACGA 676
    ::||| |||:|||||:|||||:|||||:|||||:|||||:
735 eTryIlePro.....738
675 ACCGTCACACAGTTCATTTGCTATGATGTCACGACGATTTATCATGA 626
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:
739 .....GlyLeuLeuPheAlaValAlaAsnGlyLysProIleHisPheGly 751
625 ATGGACATGCGATTTGAAACGCTGTATCTTTGCGCGGTTGTCATTT 576
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
752 .....AspGlnGluProValGlnIlyPhe.....759
575 CCGCTTTGTCGATGATGTAACGATGATGATGTAACGATGTAACGAT 526
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:
760 ....ValMetAsnPheSerAsnGlyGluIleLysPheIle.....772
525 GATGATGTCCTCTTGTAATATTTGCTAATATTAATTTGGAATATCCAAC 476
    ||| ::|||:|||||:|||||:|||||:|||||:|||||:
773 .....LysProValAlaGlyLysIlePheAspPheTronH 783
475 ACATTTAATGCGTGGCCAAACAGCTCAGATTAACAATGCGGATGAT 426
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
783 sAspIleValAlaSerGluAspGlyThrValTyrIleGlyAspAlaIleT 800
425 CACAGCTTTCTATCAATGCGACATGATATTCATTTAAGAAGCCAAAT 376
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:
800 hAsnThrValIleTrpLysPheThrLeuThrLysLeuGlnHisArgSer 816
375 AGCGAATGTTTCGACCAATGTTGAGAACCACAGATTCGAGAGCTGT 326
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:
817 ValLysLysAlaGlyIleGluValAlaGlnGluIleLysGluAlaGluAla 833
325 TAAACAGGTGTCGCGACAGAAACCTGCTGACCTGCGCAATTCGTT 276
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
833 ValGluThrIlysmetGluAsnIlysProThrSerSerGluLeuGlnLys 850
275 TACTCAGAAAGATCT.....GCAGAACCGGAGAAATATCATGTATGTA 232
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:
850 etGlnIlysglnIlyLeuIleLysGlnProGlySerGlyValProVal 866
231 CGAAGTATATCAACACCCCTTGAAATTAAGCATGATTAATCAAGCTTTGCC 182
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
867 .ValLeuIleThrThrLeuLeuValIleProValValValLeuLeuAlaI 883
181 AGTTGATTTAAGTC 168
883 |LeaIlePheIle 887

```

seq_name: A_Geneseq_36:R74171

seq_documentation_block:

```

ID R74171 standard; Protein: 3038 AA.
AC R74171;

```

```

DT 18-JAN-1996 (first entry)
DE Aspergillus terreus triol polyketide synthase.
KM Triol polyketide synthase; TPks; HMG-CoA reductase inhibitor;
KW hypercholesterolemia; LDL- cholesterol.
OS Aspergillus terreus.
FH Key Location/Qualifiers
FT region /label= keto-acyl synthase motif
FT region /label= acetyl/malonyl transferase motif
FT region 985..994
FT region /label= dehydratase motif
FT region 1446..1450
FT region /label= methyl transferase motif
FT region 1932..1937
FT region /label= enoyl reductase motif
FT region 2164..2169
FT region /label= keto reductase motif
FT region 2498
FT region /label= acyl carrier protein motif
FT misc_difference 282..288
FT misc_difference /label= misc feature
FT misc_difference 1450..1460
FT misc_difference /label= misc feature
FT misc_difference 1603..1612
FT misc_difference /label= misc feature
FT misc_difference 2521..2535
FT misc_difference /label= misc feature
PN MO9512661-A.
PD 11-MAR-1995.
PF 28-OCT-1994; U12423.
PR 02-NOV-1993; US-148132.
PA (MERI ) MERCK & CO INC.
PI Conder MJ, Davis CR, Hendrickson LE, Mcada PC, Rambosek J,
PI Reeves CD, Vinci VA;
DR WPT: 95-193816/25.
DR N-PSDB: Q92323.
PT Novel DNA encoding triol poly-ketide synthase - used to isolate and
PT identify homologues of triol poly-ketide synthase, and in the treatment
PT of hyper-cholesterolaemia
PS Claim 12: Figure 2: 107pp: English.
CC The full-length TPks-encoding DNA in plasmid pLOA was
CC designated pTKS100. Splicing of the introns from the DNA
CC sequence and translation of the 9114 nt ORF results in a
CC protein of 3038 AAs (R74171) with a mol. wt. of 269,090
CC daltons. Inspection of the TPks AA sequence for active
CC site residues and motifs known to be associated with
CC polyketide synthases and fatty acid synthase (FAS) activities
CC resulted in the identification of candidates for expected
CC sites (see FT). Except for the presence of a methyl transferase,
CC not present in FAS, the succession of activities on the
CC TPks protein is the same as that observed for the rat FAS
CC protein.
SQ Sequence 3038 AA;

alignment_scores:
    Quality: 87.00      Length: 472
    Ratio: 0.481      Gaps: 24
    Percent Similarity: 38.347      Percent Identity: 19.703

alignment_block:
us-09-323-427-5/rev x R74171 ..
Align seg 1/1 to: R74171 from: 1 to: 3038

1146 ATGGCTTTCTGACTACACTTATTGCTTGTATTGATCCGGTTGA 1097
      :::::::::::::::::::: |||
1175 ValAlaPheGlnThrValIleGlyAla.....TyrSerSerPro..... 1187
      |||::: ::::::::::::::
1096 CATGTGCTCGAAGTCAGCCAGAAATTGAATGT.....GCACCACTT 1053
      |||::: ::::::::::::::
1188 .....GlyAspArgArgLeuArgCysLeuTyrValProThrH 1200

```

```

1052 CAATA..... 1048
      :::
1200 isValaAspArgIleThrLeuValProSerIleuCysLeuAlaThrAlaGlu 1216
      :::::::::::::::::::: |||
1047 .....ACATCAATTTTAAATACAGTAATGATTCGAAGACA 1010
      :::::::::::::::::::: |||
1217 SerGlyCysGluSerValAlaPheAsnThrIleAsnThrTyrAspGly 1233
      |||::: ::::::::::::::
1009 TGTTTATGTGAAGGT.....CTTTATGATCAAGAAGTTGCCGTA 969
      |||::: ::::::::::::::
1233 yAspTyrLeuSerGlyAspIleValAlaPheAspAlaGlu..... 1246
      :::
968 ATGATGAAGCTGACGCTCAAGTTGCCGAATTCACATTCATTTGAT 922
      :::
1247 ..GlnThrThrLeuPheGlnValGluAsnIleThrPheIleProPheSer 1262
      :::::::::::::::::::: |||
921 TCATGCATGTTCGGCGTACAGCATCTCTGAATCCAGCTGATATTTT.. 874
      :::::::::::::::::::: |||
1263 ProProAspAlaSerThrAspHisAlaMetPheAlaArgTrpSerTrpG 1279
      :::::::::::::::::::: |||
873 ....GTACACACACAGCTGTTCATTTTCATCCATTTTGTATACCA 828
      :::::::::::::::::::: |||
1279 yProLeuThrProAspSerLeuAspAsnProGluTyrTrpAlaThrA 1296
      :::::::::::::::::::: |||
827 AAGTTGATCGT.....GCATATGACATCAATGCTTTTAC 793
      :::::::::::::::::::: |||
1296 IagGlnAspLysGluAlaIleProIleIleGluArgIleValTyrPheTyr 1312
      :::::::::::::::::::: |||
792 ATGGAAGCTGATAAACAAGTATGACACAGATTGAGTATCTGAATCAC 743
      :::
1313 Ile.....ArgSerPheLeuSerGlnLeuThrLeuGluGluArgG 1326
      :::::::::::::::::::: |||
742 AACTGCT.....TTTCAACTCAAAATGTGCCGATCCAGTATGCC 702
      :::::::::::::::::::: |||
1326 ngAlaAlaIaPheHisLeuGlnLysGlnIleGluTrpLeuGlu..... 1340
      :::::::::::::::::::: |||
701 GTTATGAATTTTGAATGTGACCAACCGTCACAGTTCATATTTGCT 652
      :::::::::::::::::::: |||
1341 .....GlnValLeuAlaSerAlaLysGluLysArg..... 1350
      :::::::::::::::::::: |||
651 ATCATGTGTCAGCCAGTTTATCTAATAATGACATGCGATTTCTAACCCT 602
      :::::::::::::::::::: |||
1351 .....HisLeuTrpTyrAspProGluTyrGluAs 1360
      :::::::::::::::::::: |||
601 TGATACTTTCGGCGGTGTGCTCATTCCTGCTTTGGATGAT..... 559
      :::::::::::::::::::: |||
1360 nAspThrGluAlaGlnIleGluHisLeuCysThrAlaAsnSerTyrHisP 1377
      :::::::::::::::::::: |||
559 ..... 559
1377 roHisValArgLeuValGlnArgValGlyGlnHisLeuLeuProThrVal 1393
      :::::::::::::::::::: |||
558 ...GCTACGCGTATACTGTGAAATTTCTAATGCTGATGATGCTCT 512
      :::::::::::::::::::: |||
1394 ArgSerAsnGlyAsnProPheAspLeuLeuAspHisAspGly...LeuLe 1409
      :::::::::::::::::::: |||
511 TGATTAATATTTGCTTAATTAAT.....TTGGAATATC 480
      :::::::::::::::::::: |||
1409 uThrGluPheTyrThrAsnThrLeuSerPheGlyProAlaLeuHisTyrA 1426
      :::::::::::::::::::: |||
479 CAACAGATTAAATGCTGGCCAGAAAGCTCAGTATACAAATATGGCGAT 430
      :::::::::::::::::::: |||
1426 IaArgGluLeuValAla...GlnIleAlaHisArgTyrGlnSerMetAsp 1441
      :::::::::::::::::::: |||
429 CGATTCACAGCTTTTCTATCAATGCCAGATCAGTATTAACATTAAAGACC 380
      :::::::::::::::::::: |||
1442 IleLeuGluIle..... 1445
      :::::::::::::::::::: |||
379 AAATAGCAATGTGTTGACACACATGTTCAAGACCAAGATGTGGAG 330
      :::::::::::::::::::: |||
1446 .....GlyAlaGly. 1448

```

```

329 CIGTTTAAACAGCTGTGCGCAGCAAAAACCTGCTGCAGCTGCCAATTC 280
      ||| ||||| ||||| |||||
1449 ..... ThrGlyAlaIleThrLysTYrValLeuAlaThProGlnIleu 1462
280 ..... 280

1463 GlyPheAsnSerTYrThrTYrThrAspIleSerThrGlyPhePheGluGI 1479
280 ..... 280

1479 AlaLaArgGluGlnPheAlaIarProPheGluAspArgMetValPheGluPrOL 1496
279 ..... CGTTTAACAAGAAAGATCTGCAGAACCAGAAATAATC 241
      ||| :||| :||| |::| :|
1496 euAspIleArgTrgSerProAlaGluGlnGlyPheGluPro... HisAla 1511
240 ATTGAATGTATCGAACATATATCAACACCCCTGAAATTACGGATGATATCA 191
      |||::| :|||::|::|::|::|::|::|::|::|::|::|::|::|
1512 TYrAspLeuIleIeAlaSerAsnValLeuHisAlaThrProAsp.... 1526
190 AGCTTGCCAGTTGATTAACGTCAC... CGTGCACCTGTGCAACATATG 144
1527 ... LeuGluYrSThrMetAlaHisAlaArgSerLeuLeuYSProGlyc 1542
143 GACAACTGTAATACT ..... 127
      ||||| ||||| |||||
1542 LysIleMetValIleLeuGluIleThrHisLysGluHisThrArgLeuGly 1558
126 ..... CGTCGACGTACAAAAATGGAAAT 107
1559 PheIlePheGlyLeuPheAlaAspTrpTrpAlaGlyValAspAspGlyAr 1575
106 CTGCATGTCACCATTT 91
1575 GCYSThrGluProPhe 1580
```

Quality:	86.50	Length:	359
Ratio:	0.588	Gaps:	14
Percent Similarity:	43.363	Percent Identity:	17.994

851
282 CTTGTTTACTCAGAAAGATGTCGACACCGAGATATCATTTGATG 233
852 MetGlnGluLysGlnLysLeuSerThrGluProGlySerGlyValSerVa 868
232 ACGAAGTATATCAACACCCCTGAAATTAACGATGATTAATCAAGCTTGG 183
868 L.ValLeuIleThrThrLeuLeuValIleProValIleValLeuLeuAla 884
182 CAGTTGATTACGTC 168
885 L.IleValMetPheIle 889

seq_name: A_Geneseq_36:R10323

seq_documentation_block:

ID R10323 standard; Protein: 944 AA.

AC R10323;

DT 08-APR-1991 (first entry)

DE Type B alpha-amidating enzyme.

KW Amidation: post-translational modification; AE.

OS Rattus rattus.

PN AU9049043-A.

PD 29-NOV-1990.

PF 01-FEB-1990; 049043.

PR 06-FEB-1989; US-307336.

PA (UNIG-) UNIGENE LAB INC.

PI Betelsen AH, Mehta NM, Beaudry GA;

DR WPI: 91-02243/04.

N-PSDB: Q10278.

PT Expression systems for amidating enzyme - comprises prokaryotic

or eurykaryotic hosts contg. a recombinant expression vector contg.

the corresp. DNA sequence.

PS Disclosure: Page 7; 52pp; English.

CC The sequence was derived from DNA isolated from libraries prepd.

CC from RNA extracted from rat medullary thyroid carcinoma tissues or

CC derived cell lines such as IVI 10028. The cDNAs extracted could

CC be divided into distinct types, the sequence below being encoded by

CC Type B. (Type A is given in R10322). The sequence can be inserted

CC into expression vectors for the prodn. of recombinant alpha-

CC amidating anzyme (AE) useful for post translational modification

CC of other recombinant polypeptides such as calcitonin. The sequence

CC includes a membrane spanning domain which may be undesirable in a

CC recombinant DNA expression system, possibly inactivating the enzyme.

CC A stop codon can be placed upstream of the domain-encoding sequence.

CC pref. between posns. 2340 and 2690 (counting from the CDS). This

CC results in the expression of an 87 rather than a 105 kd prod. The

CC sequences of Type A and Type B are substantially identical with the

CC exception of an intron region from bases 1178-1492 (CDS) of the

CC Type B cDNA.

SO Sequence 944 AA;

alignment_scores:

Quality: 86.50 Length: 339

Ratio: 0.588 Gaps: 14

Percent Similarity: 43.363 Percent Identity: 17.994

alignment_block:

US-09-323-427-5/rev x R10323 ..

Align seg 1/1 to: R10323 from: 1 to: 944

1065 TGTGGACCACTTCATTAACATCAATTTAATACAGTAATGATTGCA 1016
111 |||||
634 GysGlnProThrSerPvalAlaGluProSerThr..... 645
1015 AGAAGATGTTATGTGAAGCTCTTATGATCAAGAAGTTGCCGTAATG 966
111 |||||
646 G.LysAlaValPheValSerAspGlyTyr..... 654

965 ATGAAGGTGGACGTCAAGTTGCCGAATTTCACCTTCATTGATTGATGC 916
655
915 AATGTTGGCGGTACAGATCTGTGATTCACCGATGATTTTGTAAACAC 866
656 Asn...SerArgIleValGlnPheSerProSerGlyLysPheValThrGl 671
865 A.....
671 nTPGlyGluGluSerSerGlySerSerProArgProGlyGlnPheSerV 688
864ACTGTTGATTCCTTCATTCATTCATTTATTTGTTACCAAGTT 823
688 alProHisSerLeuAlaIleValAlProHisLeuAspGlnLeuCysValAla 704
822 GATCGTGCATAT...CGAGTACAATGCTTTTACATGGAAGCTGATTAAC 776
705 AspArgGluAsnGlyArgIleGlnCysPheLysThrAspThrLysGluPh 721
775 AGTTAGTGCACAGATTGAGGTATCTGAATACACAGCTGCTTCAACATC 726
721 eValArgGluIleLysHisLaserPheGlyArgAsnValPheAlaIleS 738
725 AATTTGTCGGATGCCAGTATGCCGTTATGAATTTTGGATGGAGACA 676
738 eTyrIlePro.....
675 ACCGGTCAACCACTTCATTTGCTATCATTTGTCACGCAATTCATGA 626
742GlyPheLeuPheAlaValAsnGlyLysProTyrPheGly.. 754
625 ATGACATGCGATTCGAAACCCGTGATCTTCTCGCGGTTGTCATTT 576
755AspGlnGluProValGlnGlyPhe..... 762
575 CTTGCTTTTTCGATGATGTAACGGTATGATCTGCAATTTCAATGCT 526
763ValMetAsnPheSerSerGlyLysIleLeuValPhe..... 775
525 GATGATGTCCTCTGATTAATTTGCTAATTAATTTGGAATTCACAC 476
776LysProValArgLysHisPheAspMetProHn 786
475 AGATTAAATGCTGGCCACAGACCTCAGCTATAC..... 442
786 sasPLeuValAlaSerGluAspGlyThrValTyrIleGlyAspAlaHisT 803
441AATATGCGGATGATCACAGCTTTTCTATCAATGC 406
803 hrasnThrValTyrLysPheThrLeuThrGluLysMetGlnHisArgSer 819
405CAGATCAGTATTCACATTAAAGAACCAATAGGCAATGTGT.. 364
820 ValLysLysAlaGlyIleGluValGlnGluLysAlaGluValAlaVala 836
363CGACCACAATGTTTCAGAACACACAGATTCG 333
836 GluProLysValGluAsnLysProThrSerSerGluLeuGln..... 850
332 GAGCTGTTAAACAGGTGTCGCCGACAGCAAAACCTCTGACGTCGCCGA 283
851Lys 851
282 CTTGTTTACTCAGAAAGATCTGACAAACCGAGATATCATTTGATGT 233
852 MetGlnGluLysGlnLysLeuSerThrGluProGlySerGlyValSerVa 868
232 ACGAAGTATATCAACACCCCTGAAATTAACGATGATTAATCAAGCTTGG 183
868 L.ValLeuIleThrThrLeuLeuValIleProValIleValLeuLeuAla 884
182 CAGTTGATTACGTC 168

seq_name: A_Geneseq_36:W13009

seq_documentation_block:
ID W13009 standard; protein; 560 AA.

AC W13009;
DT 21-NOV-1997 (first entry)
DE Segment of desmosomal cadherin, desmoglein Dsg2.
KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;
KW carcinoma; desmosome; antibody; epitope; diagnosis; detection;
KW micrometastasis; separation; enrichment; targeted delivery;
KW metastatic.
OS Homo sapiens.
PN DE19531033-A1.
PD 27-FEB-1997.
PF 23-AUG-1995; 031033.
PR 23-AUG-1995; DE-031033.
PA (PROG-) PROGEN BIOTECHNIK GMBH.
PI Franke WW, Schaefer S;
DR WPI: 97-146518/14.
PT Antibody reactive with part of desmosomal cadherin - exposed on
PT surface of epithelial or carcinoma cells, not bound to desmosomes,
PT useful for diagnosis and treatment of carcinoma micrometastases
PS Claim 7; Page 5; 8pp; German.
CC The present sequence is a segment of the desmosomal cadherin (DC),
CC desmoglein Dsg2, which is exposed on the surface of epithelial or
CC carcinoma cells and not bound to desmosomes. An antibody (Ab)
CC directed against epitopes of the present sequence can be used to
CC diagnose, i.e. to detect carcinoma cells, especially
CC micrometastases, not bound to desmosomes, to separate, enrich or
CC detect living or fixed carcinoma cells by cell sorting methods and
CC as a therapeutic to deliver agents, e.g. other Ab or toxins, to
CC target cells. The Ab provides rapid and reliable detection of
CC metastatic carcinoma, and detects parts of DC that are not
CC accessible in desmosome bound cells, as in normal tissue or
CC carcinomas.
SQ Sequence 560 AA:

alignment_scores:
Quality: 85.50 Length: 203
Ratio: 0.822 Gaps: 12
Percent Similarity: 51.232 Percent Identity: 25.616

alignment_block:
US-09-323-427-5/rev x W13009 ..

Align seg 1/1 to: W13009 from: 1 to: 560

```

1119 TTGTCTTATTCATCCGGTGCATGTCGCAAGGTCAGCCAGAAAT 1070
      |||:::|||||:::|||||:::|||||:::|||||:::
299 LLeaSpHeSeRValIleValAlaAsnLysAlaAlaPheHisAspSerI 315
      |||:::|||||:::|||||:::|||||:::|||||:::
1069 T.....GAATGTGCACCACTTCATATCAATCAATTTTATACCGTA 1026
      |:::|||||:::|||||:::|||||:::|||||:::
315 eArgeSerLysTyrLysProThrProLeuProLeuValLysValLysA 332
      |||:::|||||:::|||||:::|||||:::|||||:::
1025 ATGCATTCGAGGA.....CATGTTTATGTG 1000
      |||:::|||||:::|||||:::|||||:::|||||:::
332 snValLysgluLylIleHisPheLysSerSerValIleSerIleTyrVal 348
      |||:::|||||:::|||||:::|||||:::|||||:::
999 AAAGGCTTTATGATCAAGAAGTCCCTAATGATGAGGTCGACGTCA 950
      :::|||||:::|||||:::|||||:::|||||:::
349 SerGlusSerMetAsp.....ArgSerSerLysGly.....GI 359
      |||:::|||||:::|||||:::|||||:::|||||:::
949 ACTTGCAGCAATTTCACTTCATTCATTCATTCATTCATTCATTCAT 912
      |:::|||||:::|||||:::|||||:::|||||:::
359 nIleIleIleGlyAsnPhcGlnAlaPheAspGlnAspThrGlyLeuProAlaH 376
      |||:::|||||:::|||||:::|||||:::|||||:::
911 TTGCGCGTACACGATCTGGAATCCACGTGTATTTT..... 874
      |||:::|||||:::|||||:::|||||:::|||||:::

```

```

376 IsAlaArgTyrValLysLeuGluAspArgAspAsnTrpIleSerValAsp 392
      |||:::|||||:::|||||:::|||||:::|||||:::
873 ...GTACACAACTGCTTCATTTGCTTCATCCATTTATTTGTCACAA 827
      |||:::|||||:::|||||:::|||||:::|||||:::
393 SerValThrSerGluIleLysLeuAlaLysLeuProAspPheIleUserAr 409
      |||:::|||||:::|||||:::|||||:::|||||:::
826 AGTT.....GATCGCATATCGATGACATGCTTTTACATGAGACGTG 783
      |:::|||||:::|||||:::|||||:::|||||:::
409 gTyrValGlnAsnGlyThrTyrThrValLysIleValAlaIleSerGlnA 426
      |||:::|||||:::|||||:::|||||:::|||||:::
782 AT.....AAACAGTTAGTGACAG.....ATTGAGGTATTCGAA 748
      |||:::|||||:::|||||:::|||||:::|||||:::
426 sPTyrProArgLysThrIleThrGlyThrValLeuIleAsnValGluAsp 442
      |||:::|||||:::|||||:::|||||:::|||||:::
747 ATCACACACTGCTTTTCAACTCAATTCATTCGCGATG...CCAGTATCCG 701
      |||:::|||||:::|||||:::|||||:::|||||:::
443 IleAsnAspAsnCysProThrLeuIleGluProValGlnThrIleCysH 459
      |||:::|||||:::|||||:::|||||:::|||||:::
700 T.....TATGAAATTTTGATGGTGACCA 675
      |||:::|||||:::|||||:::|||||:::|||||:::
459 sAspAlaGluTyrValAsnValThrAlaGluAspLeuAspGlyHisProA 476
      |||:::|||||:::|||||:::|||||:::|||||:::
674 CCGGTCAACCACTTCATTTGCTATTCATTCGTCACCA.....GTTTAT 631
      |||:::|||||:::|||||:::|||||:::|||||:::
476 snSerGlyProPheSerPheSerValIleAspLysProProGlyMetAla 492
      |||:::|||||:::|||||:::|||||:::|||||:::
630 CATTAATGG 622
      |||:::|||||:::|||||:::|||||:::|||||:::
493 GluLysTrp 495

```

seq_name: A_Geneseq_36:R20026

seq_documentation_block:
ID R20026 standard; Protein; 973 AA.

AC R20026;
DT 08-APR-1992 (first entry)
DE C-terminal amidation enzyme.
KW Horse.
OS Equus caballus.
FH Key
FT Location/Qualifiers
FT Region 829..895
FT /note="deleted in R20027 and R20028"
FT
FT J03262484-A.
PD 22-NOV-1991.
PF 14-MAR-1990; 063306.
PR 14-MAR-1990; JP-063306.
PA (SHIS) SHISEIDO KK.
DR WPI: 92-012701/02.
DR N-PSDB; Q20198.
PT cDNA sequence - coding for peptide C terminal amidation enzyme
PT obtd. from horse.
PS Claim 1; Fig 1; 18pp; Japanese.
CC The sequence was deduced from the DNA sequence determined from a
CC clone isolated from a library prepd. from mRNA extracted from
CC horse atrium tissue. The sequence is one of four similar ones
CC provided which all have the same sequence up to residue 809 at
CC which point they diverge, having different C-terminal. The
CC different termini are created by deletions in the last portion of
CC DNA (see feature table).
CC See also R20025-28.
SQ Sequence 973 AA:

alignment_scores:
Quality: 85.00 Length: 332
Ratio: 0.582 Gaps: 14
Percent Similarity: 43.976 Percent Identity: 18.072

alignment_block:
US-09-323-427-5/rev x R20026 ..

Align seg 1/1 to: R20026 from: 1 to: 973

```
1065 TGTGACCACTTCATTAACAAATCAATTTTAAATACAGTAATGCATTCGA 1016
    ||| |||||::: ||| |||||::: |||
630 CysGlnProThrAspValAlaValAspProAsnThr..... 641
1015 AGGACATGTTTATGTCAAGGCTTTATGATCAAGAAGGTGCCGTAAG 966
    ||| ::::: ||| |||
642 .GlyThrIlePheValSerAspGlyTyr..... 650
965 ATGAAGGTGACGTCAGATTGCCGAATTTCACTTCATTTGATTCATCGC 916
    ||| ..... Cys 651
651 .....
915 AATGTGCGGTACAGATCTGAAATCCAGGTATTTTGTAAACAC 866
    ||| ::::: ||| ||| ||::: |||
652 Asn...SerArgIleValGlnPheSerProThrGlyArgPheIleThrGI 667
865 A..... 865
667 nTrpGlyGlnGluSerSerGluSerAsnProLysProGlyGlnPheArgV 684
864 .....ACTGTTGTCATTTGCTTTCATCCATTTATTTGTTACCAAGTT 823
    ::::: ||| ::::: |||
684 aLProHisSerLeuAlaLeuValProHisLeuGlyGlnLeuCysValAla 700
822 GATCGTGCATAT...CGAGTACAAATGCTTTTACATGGAAGCTGATAAAC 776
    ||||| ||::: ||||| ||::: |||
701 AspArgGluAsnGlyArgIleGlnCysPheLysThrAspThrLysGluPh 717
775 AGTTATGTCACAGATGAGGTATCTGAATCAACTGCTTTTCAAACTC 726
    ||| ::::: ||| ::::: |||
717 eValArgGluIleLysHisAlaSerPheGlyArgAsnValPheAlaIleS 734
725 AAATGTCCCGATGCCAGTATGCCGTTATGAATTTTGATGCTGGACCA 676
    ::: |||||
734 eTyrtIlePro..... 737
675 ACCGGTCAACCACTCAATTGCTATCATTTGTCACGCACTTATCATAA 626
    ::: |||||::: ||::: ||| :::
738 .....GlyLeuLeuPheAlaValAsnGlyLysProGlyPheGly... 750
625 ATGGACATGCGATTCTGAAACCGTTGATACTTTCGCCGCGTGTCCATT 576
    ||::: |||
751 .....AspGlnLysProValGlnGlyPhe..... 758
575 CCGCTTTGCGATGATGATGACGATGATCTGGAATTTGTAATGCT 526
    ::::: |||||::: ||::: ||| :::
759 .....ValMetAsnPheSerSerGlyGluIleLeuAspValPhe..... 771
525 GATGATGTGCTCTTGATAAATATTGCTAAATATTGGAATATCCAAC 476
    || ::: |||
772 .....LysProValArgLysHisPheAspMetProH 782
475 AGATTTAATGCGTGGCCAGAAAGCTCAGCATAC..... 442
    ||::: |||||::: |||||
782 sAspIleThrAlaSerGluAspGlyThrValGlyValGlyAspAlaHisT 799
441 .....AAATATGCGATGATCAACAGCTTTTCTATCAATGC 406
    ||::: |||||::: |||||
799 hTrAsnThrValTrpLysPheThrSerThrGluArgValGluHisArg... 814
405 CAGATCAGATTATACATTAAGAACCAAAATAGCGAATGTGTCGACCACA 356
    ::::: |||||::: ||| :::
815 .....SerValLysLysAlaGlyIleGluValGlnGluIleLys 827
355 ATGTTACAGAACCAAGATTCGAGCTGTTAAACAGAGGTGCGCCGAC 306
    ::: |||||
827 sGluSerGlu.....AlaValValGluThrLysMetGluA 839
305 CAAAACCTGCTGCGCAACTGCTGTTTACTCAAGAAAGATCT... 259
    ||||| ||::: |||||::: |||
839 snLysProAlaSerSerGluLeuGlnLysMetGlnLysGlnLysLeu 855
258 ...GCAGAACCGAGAAATATCATTTGATGTAAGAACTGATATCAACACCT 212
```

```
||||| ::: ||| |||||::: |||
856 IleLysGluProGlySerGlyValProValValLeuIleThrThrLeu 872
211 TGAATTTAGCGATGATTAATCAAGCTTTGCCAGTTGATTACGTC 168
    || ::: |||||::: |||
872 euValIleProValValValLeuLeuAlaIleAlaIlePheIle 886
```

THIS PAGE BLANK (USPTO)

OM of: US-09-323-427-5 to: Issued_Patents_AA:* out_format : pfs
Date: Apr 16, 2000 4:46 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+ndp.model -DEV=xlp
-Q=/cgn2_1/USPRO.spool/US09323427/runat_14042000_170513_19867/app-query.fasta.1
-DB=Issued_Patents_AA -QFMT=fastx -SUFFIX=backtrans.ra1
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000
-LOOPEXT=0.000 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -XGAPOP=6.000 -XGAPEXT=7.000 -XGAPOP=10.000
-XGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blolsum62 -TRANS=human40.cdi -LIST=45 -DOCLIGN=200
-THR_SCORE=pct -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
-MINLEN=0 -MAXLEN=1000000 -USER=US09323427 -NCPU=6 -ICPU=3
-NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-323-427-5
Query length: 1161
Database: Issued_Patents_AA:*
Database sequences: 13390
Database length: 1329546
Search time (sec): 99.320000

score_list:

Sequence	Strd Orig	Zscore	EScore	Len	! Documentation	..	935	1	5
/cgn2_6/pctoddata/1/iaa/5A.COMB.pcp:US-07-707-367-2		100.00	163.44	0.0075	989	1	989	1	5
/cgn2_6/pctoddata/1/iaa/5B.COMB.pcp:US-08-070-301-16		96.50	155.51	0.0195	300	1	300	1	5
/cgn2_6/pctoddata/1/iaa/5C.COMB.pcp:US-08-477-451-8		96.50	141.07	0.0385	801	1	801	1	5
/cgn2_6/pctoddata/1/iaa/5D.COMB.pcp:US-07-906-349A-6		93.00	150.86	0.0438	362	1	362	1	5
/cgn2_6/pctoddata/1/iaa/5E.COMB.pcp:US-08-415-751-35		91.50	157.52	0.0418	650	1	650	1	5
/cgn2_6/pctoddata/1/iaa/5F.COMB.pcp:US-08-325-071-67		90.00	147.22	0.0861	549	1	549	1	5
/cgn2_6/pctoddata/1/iaa/5G.COMB.pcp:US-08-325-071-61		89.00	147.23	0.1018	650	1	650	1	5
/cgn2_6/pctoddata/1/iaa/5H.COMB.pcp:US-08-325-071-63		89.00	145.15	0.1123	650	1	650	1	5
/cgn2_6/pctoddata/1/iaa/5I.COMB.pcp:US-08-450-332-2		87.00	122.05	0.4648	3038	1	3038	1	5
/cgn2_6/pctoddata/1/iaa/5J.COMB.pcp:US-08-437-640-2		87.00	122.05	0.4648	3038	1	3038	1	5
/cgn2_6/pctoddata/1/iaa/5K.COMB.pcp:US-08-440-804-12		85.50	130.11	0.4099	1226	1	1226	1	5
/cgn2_6/pctoddata/1/iaa/5L.COMB.pcp:US-08-218-265-12		85.50	130.11	0.4099	1226	1	1226	1	5
/cgn2_6/pctoddata/1/iaa/5M.COMB.pcp:US-08-480-604A-6		85.00	119.32	0.7398	2710	1	2710	1	5
/cgn2_6/pctoddata/1/iaa/5N.COMB.pcp:US-08-325-071-65		85.00	119.32	0.7398	2710	1	2710	1	5
/cgn2_6/pctoddata/1/iaa/5O.COMB.pcp:US-08-405-496A-6		84.00	135.39	0.4118	620	1	620	1	5
/cgn2_6/pctoddata/1/iaa/5P.COMB.pcp:US-08-070-301-14		83.50	128.61	0.6157	989	1	989	1	5
/cgn2_6/pctoddata/1/iaa/5Q.COMB.pcp:US-08-682-517-15		83.00	124.98	0.7944	1222	1	1222	1	5
/cgn2_6/pctoddata/1/iaa/5R.COMB.pcp:US-08-682-517-9		83.00	124.68	0.8056	1252	1	1252	1	5
/cgn2_6/pctoddata/1/iaa/5S.COMB.pcp:US-08-162-081B-36		82.00	124.43	0.9645	1080	1	1080	1	5
/cgn2_6/pctoddata/1/iaa/5T.COMB.pcp:US-08-780-872-36		82.00	124.43	0.9645	1080	1	1080	1	5
/cgn2_6/pctoddata/1/iaa/5U.COMB.pcp:US-08-348-006B-5		81.50	116.37	1.53	1911	1	1911	1	5
/cgn2_6/pctoddata/1/iaa/5V.COMB.pcp:US-08-800-825A-5		81.50	116.37	1.53	1911	1	1911	1	5
/cgn2_6/pctoddata/1/iaa/5W.COMB.pcp:PCP-US94-10166-5		81.50	116.37	1.53	1911	1	1911	1	5
/cgn2_6/pctoddata/1/iaa/5X.COMB.pcp:PCP-US94-10166-5		81.50	110.13	2.05	3174	1	3174	1	5
/cgn2_6/pctoddata/1/iaa/5Y.COMB.pcp:US-08-477-451-3		81.00	134.36	0.7161	407	1	407	1	5
/cgn2_6/pctoddata/1/iaa/5Z.COMB.pcp:US-08-162-040A-3		81.00	122.48	1.25	1069	1	1069	1	5
/cgn2_6/pctoddata/1/iaa/5A.COMB.pcp:US-08-780-872-37		80.50	122.48	1.25	1069	1	1069	1	5
/cgn2_6/pctoddata/1/iaa/5B.COMB.pcp:US-08-415-751-6		80.50	134.74	0.7643	362	1	362	1	5
/cgn2_6/pctoddata/1/iaa/5C.COMB.pcp:US-08-785-076-3		80.50	132.65	0.8441	430	1	430	1	5
/cgn2_6/pctoddata/1/iaa/5D.COMB.pcp:US-08-785-076-2		80.50	132.65	0.8464	432	1	432	1	5
/cgn2_6/pctoddata/1/iaa/5E.COMB.pcp:US-08-325-071-59		80.00	126.53	1.22	650	1	650	1	5
/cgn2_6/pctoddata/1/iaa/5F.COMB.pcp:US-08-484-993B-16		79.00	126.08	1.48	570	1	570	1	5
/cgn2_6/pctoddata/1/iaa/5G.COMB.pcp:US-08-484-158B-16		79.00	126.08	1.48	570	1	570	1	5
/cgn2_6/pctoddata/1/iaa/5H.COMB.pcp:US-08-484-158B-16		79.00	126.08	1.48	570	1	570	1	5
/cgn2_6/pctoddata/1/iaa/5I.COMB.pcp:US-08-480-150A-16		79.00	126.08	1.48	570	1	570	1	5
/cgn2_6/pctoddata/1/iaa/5J.COMB.pcp:US-08-458-731-16		79.00	126.08	1.48	570	1	570	1	5
/cgn2_6/pctoddata/1/iaa/5K.COMB.pcp:US-08-070-301-17		79.00	119.39	2.03	989	1	989	1	5
/cgn2_6/pctoddata/1/iaa/5L.COMB.pcp:US-08-477-459-20		78.50	128.39	1.44	434	1	434	1	5
/cgn2_6/pctoddata/1/iaa/5M.COMB.pcp:US-08-477-459-20		78.50	128.39	1.44	434	1	434	1	5
/cgn2_6/pctoddata/1/iaa/5N.COMB.pcp:PCP-US94-01826A-20		78.50	128.39	1.44	434	1	434	1	5
/cgn2_6/pctoddata/1/iaa/5O.COMB.pcp:PCP-US94-01826A-20		78.50	128.39	1.44	434	1	434	1	5
/cgn2_6/pctoddata/1/iaa/5P.COMB.pcp:PCP-US94-02252A-20		78.50	128.39	1.44	434	1	434	1	5

seq_name: /cgn2_6/pctoddata/1/iaa/5A.COMB.pcp:US-07-707-367-2

seq_documentation block:

Sequence 2, Application US/07707367
Patent No. 5196316
GENERAL INFORMATION:
APPLICANT: Iwasaki, Yasuno
APPLICANT: Shimoi, Hiroko
APPLICANT: Suzuki, Kenji
APPLICANT: Chisaba, Oreste
APPLICANT: Nishikawa, Yoshiki
APPLICANT: Kawahara, Takashi
APPLICANT: Kanagawa, Kenji
TITLE OF INVENTION: No. 5196316el Enzyme and DNA Coding Therefor
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/707,367
FILING DATE: 19910530
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 141678/90
FILING DATE: 01-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 210535/90
FILING DATE: 10-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 329911/90
FILING DATE: 30-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Villamizar, Joann
REGISTRATION NUMBER: 30,598
REFERENCE/DOCKET NUMBER: 4-18110/A/CGJ 44
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914)785-7120
TELEFAX: (914)541-8689
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 935 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-707-367-2

alignment_scores:
Quality: 100.00 Length: 391
Ratio: 0.541 Gaps: 18
Percent Similarity: 47.315 Percent Identity: 19.437

alignment_block:
US-09-323-427-5/rev x US-07-707-367-2 ..
Align seg 1/1 to: US-07-707-367-2 from: 1 to: 935
1065 TGTGGACCACTTCATTAACATTTTATACACGTAATGCA 1016
|||||

```

526 CysGlnProThrAspValAlaValAspProIleThrGlyAsnPhePheVal 542
1015 AGCAGATGTTTATGTGAAGGCTTTATGATCAGAGAAGGTGCCGTAATG 966
542 LAlaAspGlyTyr..... 546
965 ATGAAGGTGGACGTCAGAGTTGCCGGAATTCCTCCATTGATGATTCATGC 916
547 .....Cys 547
915 AATGTGGCGGTACAGATCTGAAATCCAGCGTATTTTGTGA..... 871
548 Asn...SerArgIleMetGlnPheSerProAsnGlyMetPheIleMetGln 563
871 ..... 871
563 nTrpGlyGlnGluThrSerSerAsnValProArgProGlyGlnPheArgI 580
870 .....ACAACAACGTGTTCATTTCGTTTCATCCATTATTTGTTACC 829
580 LeProHisSerLeuThrMetValProAspGlnGlyGlnLeucysValAla 596
828 AAGATTGATCGTCGATATCGAGTACATGCTTTTACATGAGCGCATATA 779
597 AspArgGlnAsnGly...ArgIleGlnCysPheHisAlaGluThrGlyAs 612
778 AACAGTTAGTCGACAGATGAGTATCTGAATC...ACAACGTGCTTTTC 732
612 nPheValIys...GlnIleIysHisGlnGluPheGlyArgGluValAlaPhe 628
731 AAATCGAAATTTGTCGCGATGCCAGTATGCCGTTTGAATTTGGATGCT 682
628 LAlaValSerTyrAlaPro.....Gly 634
681 GGACCAACCGGTCAACCAAGTTCAATTTGCTATCATGTCAGCCAGTTTA 632
635 Gly.....ValLeuTyrAlaValAsnGlyLysProIleTyr 646
631 TCATTAATGACATGCGATTCTGAACCGCTTGATACTTTTCGCGCGTTG 582
646 r.....GlyTyrSerAlaProV 652
581 TCCATTCCGCTTGTGCGATGATGTAAGGTGATACCTGTGGAATTCGA 532
652 aGlnGlnIlyPheMetLeuAsnPheSerAsnGlyAsp..... 663
531 AATGCTGATGATGTCGCTCTTGATAAATATTG.....CTAATTAATT 488
664 .....IleLeuAspThrPheIleProAlaArgLysAsnPh 675
487 GGAATATCCACAGATTTTAATGCTGGCCAGAGAAGCTCAGTATACAAT 438
675 eAspMetProHisAspIleAlaAlaAspAspGlyThrValIysValG 692
437 ATGCCGATGATCAGAGCTTTCTATCAATGCCAGATCGT..... 397
692 LysAlaPheHisAlaAsnAlaValITrPlysPheSerProSerIysAlaGlu 708
396 ...ATTACATTAAAGAACCAAT.....AGCGATGTGTCGACG 359
709 HisArgSerValLysLysAlaGlyIleGluValGlnGluIleThrGluTh 725
358 ACAATGTTTCAGAACCAAGATTCGAGCTGTTAACAGAGGTGTGCGG 309
725 rGluIlePheGluThrHisIleArgSerArgProLysThrAsnGlnSerV 742
308 CAGCAAAACCTGCTGCAGCTGCGCAACTTCGTTTACTCAGAAAGATCT 259
742 aGlnIlyLysGlnThrGlnIlyLysGlnIlyLys.....GlnIysAsnSer 756
258 GCA.....GAACCGAGATATCATTT...GATGTACGAGAC 227
757 AlaGlyValSerThrGlnGlnIlyLysGlnAsnValValGlnGlnIleAsnAl 773

```

```

226 TGATATCAACACCTTGAA.....ATTAGCGATGATTAATCAAGCTT 186
773 aGlyValAlaProThrGlnGlnIlyLysGlnAsnValValGlnGlnIlyLys 790
185 TCCGAGTTGATTTACGTACCGCTGCACCTTCTGCAACATTAATGACACCT 136
790 LysValSerThrGlnIlyLysGlnSerValValGlnGlnIlyLysSerIle 806
135 GTAATACCTTGCTGCAGTACCAAAATGAATTCGATGTCACCATTTGGCTT 86
807 ValSerThrGlnGlnIlyLysGlnSerValValGlnGlnIlyLysSerIle 823
85 CTCAAATGTTTATGGGTTTAAGCATTCATTCATTT...GCTGCCGTCAATTA 39
823 LserPheValIleuIleIleThrLeuIleuIleIleProIleAlaValIleu 840
38 TTACCATTTCCGTTTAATTTGCT 16
840 LeAlaIleAlaIlePheIleArg 847

seq_name: /cgn2_6/ptodata/1/laa/5b_cowb.pep:us-08-070-301-16
seq_documentation_block:
; Sequence 16, Application US/08070301
; Patent No. 5871995
; GENERAL INFORMATION:
; APPLICANT: IIDA, Toshio
; APPLICANT: KAMINUMA, Toshihiko
; APPLICANT: FUSE, Yuka
; APPLICANT: TAJIMA, Masahiro
; APPLICANT: YANAGI, Mitsuo
; APPLICANT: OKAMOTO, Hiroshi
; APPLICANT: KISHIMOTO, Jiro
; APPLICANT: IFUKU, Ohji
; APPLICANT: KATO, Ichiro
; TITLE OF INVENTION: ENZYME PARTICIPATING IN C-TERMINAL
; TITLE OF INVENTION: AMIDATION, AND METHOD OF PREPARING SAME AND USE THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wegner, Cantor, Mueller & Player, P.C.
; STREET: 1233 20th Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-8218
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/070,301
; FILING DATE: 24-MAY-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1-209687
; FILING DATE: 15-AUG-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1-181933
; FILING DATE: 31-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-76331
; FILING DATE: 26-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-106412
; FILING DATE: 24-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-205475
; FILING DATE: 02-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Player, William E.
; REGISTRATION NUMBER: 31,409

```

```

; REFERENCE/DOCKET NUMBER: P-450-22830
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-040
; TELEFAX: (202) 835-0605
; TELEX: 440706
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 989 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Frog
; US-08-070-301-16

alignment_scores:
  Quality: 96.50      Length: 232
  Ratio: 0.965       Gaps: 12
  Percent Similarity: 43.103   Percent Identity: 21.552

alignment_block:
  US-09-323-427-5/rev x US-08-070-301-16 ...

Align seg 1/1 to: US-08-070-301-16 from: 1 to: 989

1065 TGTGGACCACTTCATACATCAATTTATATCCAGCTATGCTTGA 1016
    ||| |||||.....:||||| ||| |||
642 CyslnPntrhspvalalavalasproilethrglyasnphpehva 658
1015 AGGACATTTATGTGAAGGCTTTATGATCAAGAAAGTTGCCGTAATG 966
    :: |||
658 lAlasplyTyr.....:..... 662
965 ATGAGGTGGACGTCACATGCCGAATTCACCTTCATTGATTCATGC 916
663 .....:.....Cys 663
915 AATGTTGGCGGTACACGATCTCTGAAATCCAGGTGATTTTGTGA.... 871
    ||| |||||.....:||||| ||| |||
664 Asn...SerArgllemetcInpheserProasnnglymetPheIleMerGl 679
871 .....:..... 871
679 ntrpOlygluGluThrserSeranLeuProArpProglInpheArgI 696
870 .....ACAACTGTTGTTGTCATTTGTTTATTTTATTTTACC 829
    :: |||||.....: ||| |||
696 lProHIsSerleuthrmetIleSerAspGlnGlyInleucysValAla 712
828 AAGTTGATCGCATATCGAGTATGCAATGCTTTTTCATGGAAGCATGA 779
    :|||.....: |||||.....:
713 AsparGluasnGly..ArglleGlnCysPheHIsAlaLysThrGlyGl 728
778 AACAGTTAGTGCACAGATTGAGGATATCGAAATC...ACAACTGCTTTTC 732
    : |||||.....: |||||.....:
728 upheValLys...GlnlleLysHIsGlnGluPheGlyArgGluValPhea 744
731 AAACGCAATTTGTCGGATGCCAGATATGCCGTTATGAATTTTGGATGCT 682
    :|||.....: |||
744 lAlaIserTyrAlaPro.....Gly 750
681 GGACCAACCGGTCAACAGCTTCATTTGCTATTCATTTGTCAGCCGTTTA 632
    ||| |||||.....: |||||.....:
751 Gly.....ValleuTyrAlaValasnGlyLysProTyrTly 762
631 TCATTAATGACATGCGATTCTGAAACCGGTGATCTTCTGCGCGGTG 582
    : |||||.....: |||
762 rGly.....AspserThrProValGlnGlyPhe..... 771
581 TCATTTCTGCTTTGTCGATGATGTAACGGTACTGCTGGAATTTCTA 532
    :|||.....: |||||.....:

```

```

772 .....MetIleasnPheserAsnGlyAsp..... 779
531 AATGCTGATGATGTCCTCTGATTAATATTG.....CTAATAATT 488
780 .....IleLeuAspThrPheIleProAlaArgLysAsnPh 791
487 GGAATATCCACAGATTATATGGCTGGCCAGAAAGCTCAGCTATAC 442
    :||| ||| |||||.....: |||||
791 eGluMetProHIsAspIleAlaIaGlyAspAspGlyThrValTyr 806

seq.name: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:US-08-477-451-8

seq_documentation_block:
; Sequence 8, Application US/08477451
; Patent No. 5928865
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori CagI Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,451
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0335.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2708
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3200 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-477-451-8

alignment_scores:
  Quality: 96.50      Length: 369
  Ratio: 0.603       Gaps: 20
  Percent Similarity: 43.360   Percent Identity: 22.764

alignment_block:
  US-09-323-427-5 x US-08-477-451-8 ..

Align seg 1/1 to: US-08-477-451-8 from: 1 to: 3200

112 ATTTTGTACTGCAGCAAGTAT.....TACAGGTGTGCATTATGTT 152
    |||||.....: |||||.....:
745 lIleValPheCysArgllePheGluLeuValTyrGluLeuIleAlaLys 761
153 GCAGAGTGCACGGGTGACGTAATCACTGGCAAGCTTGATTATCATCG 202
    :||| ||| |||||.....: |||||
761 slystrhCys.....PheSerPheTyrArgleuIleGluIleV 774
203 CTAATTTCAAGCGTGTGATATC.....ACTTCGATCATCAATGATATT 246
    :||| ||| |||||.....: |||||
774 alaSnPheLeuSerGlnPheLeuIleProLysProTyrLeuThrGluIle 790

```

```

seq_name: /cgrn2_6/ptodate/1/iaa/5A_COMB.pep:US-07-906-349A-6
seq_documentation_block:
: Sequence 6, Application US/07906349A
: Patent No. 5434064
: GENERAL INFORMATION:
: APPLICANT: Schlessinger, Joseph
: APPLICANT: Skolnik, Edward Y.
: APPLICANT: Margolis, Benjamin I.
: TITLE OF INVENTION: A NOVEL EXPRESSION-CLONING METHOD FOR
: TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE KINASES
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Broadway and Nelmark
: STREET: 419 Seventh Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/906,349A
: FILING DATE: 30-JUN-1992
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/643,237
: FILING DATE: 18-JAN-1991
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-628-5197
: TELEFAX: 202-737-3528
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 801 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
:
: US-07-906-349A-6

alignment_scores:
Quality: 93.00 Length: 77
Ratio: 3.000 Gaps: 1
Percent Similarity: 40.260 Percent Identity: 28.571

alignment_block:
US-09-323-427-5/rev x US-07-906-349A-6 ..

Align seg 1/1 to: US-07-906-349A-6 from: 1 to: 801

346 ACCAAGAGATTGCGAGCTGTTAAACAGGTGTCGCCGACGAAACTG 297
||||| :||| ||| :|||
573 ThrTrrAlaAlaAlaThrCysAlaCysThrGlyCysThrGlyCysThrcy 589
CGTCAGCTGCGCAACTTCGTTTACTCACAAGAAAGATCGCAGAACCGAG 247
|||||:|||||:||||| ||| |||
589 scysThrCysThrThrThrCysCysCysCysThrCysCysThrThr.... 604
246 AATATCATTTGATGTACGAATGATATCAACACCCCTTGAATTTAGCGATGA 197

```

```

604 ..... 604
196 TAATCAAGCTTTCAGTGTATTACGTACCGTCGACCTTCGACACATA 147
    |||::: |||||:::||||
605 .....ThnGlyThrCysThrThrThrThr 613
146 ATGGACACCTGTAATACCTTCGTGACATACA 116
    ||||| |||||:::||||
613 rThrThrThrcysAlaThrCysCysThrThr 623.
seq_name: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:US-08-415-751-35

```

```

seq_documentation_block:
; Sequence 35, Application US/08415751
; Patent No. 5643772
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: POLYPEPTIDES BINDING ANTI-
; TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA
; TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID
; TITLE OF INVENTION: VECTOR AND TRANSFORMED HOST AND
; TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PHILLIPS, MOORE, LEMPJO & FINLEY
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: California
; COUNTRY: United States of America
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette - 3.5 inch, 1.44 KB storage
; COMPUTER: PC
; OPERATING SYSTEM: DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/415,751
; FILING DATE: 03-APR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/071,880
; FILING DATE: June 1, 1993
; APPLICATION NUMBER: 07/891,301
; FILING DATE: May 29, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hana Dolezalova
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.19-2 (HHD)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-1677
; TELEFAX: (415) 324-1678
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 362 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Cryptosporidium parvum
; FEATURE:
; NAME/KEY: Positions coded by nonsense codons are
; NAME/KEY: Identified as Xaa.
US-08-415-751-35

```

```

alignment_scores:
Quality: 91.50 Length: 151
Ratio: 1.220 Gaps: 9

```

Percent Similarity: 49.669 Percent Identity: 25.828

alignment_block:

US-09-323-427-5/rev x US-08-415-751-35 ..

Align seg 1/1 to: US-08-415-751-35 from: 1 to: 362

```

401 TCAGTATTACCATTAAGAACCAATAGCAATGTTGACAC...AA 355
    || || |||:::||||: |||::: |||::: |||
20 SerAsnLeuLeuLeuGlnHisSerArgArgHisPhePheGlnArgLeuG1 36
354 TGTTCAGAACCCACACAGAGATTGCGAGCTGTTAAACAGGTGGTCCGAC 305
    :::::|||| ||||| |||::: |||
36 yMeTArGAsn**ProHisArgGluLeuLeuHisGluValValIyG 53
304 AAACCTGCTGCAGCTGCCGAC.....TTC 279
    ||::: |||::: |||
53 InThrIle**HisLeuLysAsnArgCysLysTyrGlnSerThrHis 69
278 GTTACTCAAGAAAGAT.....CTGACAGACCGGAGAAATCATTTGAT 235
    :: |||:: ||| ||||| |||
70 LeuGlnSerGlnIleAspPheLysLeuGlnAsnArgLeu..... 82
234 GTACGACTGATATCAACACCTTGAAATTAGCGATGATATCAAGCTTT 185
    |||::: |||::: |||::: |||
83 .....LeuValGlu****LeuGlnLeuArgLeuValValLeuLeu* 97
184 GCCAGTTGATTTACGTCACCGTCGACTTTCGACACATAATGACA..... 140
    || || |::: |||::: |||::: |||
97 ****LeuGlnPhe.AspleuGValLeu*****Tyr**ArgLe 113
139 .ACCTGTAATACCTTGCTGCAGTACAAATGAGAT..... 107
    || || ||||| |||
113 ucYscYscYsglyCysCysSerIySlleTrrPaspanCysCysSel 130
106 .....CTGCATGTCACCATTTGGCTTCTC 83
    |||:: |||:: |||
130 eupHeTrrPcYscYscYscYscYsLeuAsnTrrPheLeuTrrP... 145
82 AATGTTTATGCGTTTAAGCATTTGCATTGTCGCCGTCATTATACCA 33
    ||||| ||| ||| ||| |||
146 .....TyrGly.....CysCysCysCysCys**LeuTrrPyrAs 157
32 T 32
157 n 157
seq_name: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:US-08-325-071-67
seq_documentation_block:
; Sequence 67, Application US/08325071
; Patent No. 5587311
; GENERAL INFORMATION:
; APPLICANT: COBON, Stewart Gary
; APPLICANT: MOORE, Joanna Terry
; APPLICANT: JOHNSON, Law Anthony York
; APPLICANT: WILLADSEN, Peter
; APPLICANT: KEMP, David Harold
; APPLICANT: SRISKANTHA, Alagacone
; APPLICANT: RIDING, George Alfred
; APPLICANT: RAND, Keith No. 5587311man
; TITLE OF INVENTION: DNA Encoding A Cell Membrane
; TITLE OF INVENTION: Glycoprotein of A Tick Gut
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.25
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/325,071
      FILING DATE: 14-OCT-1993
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 08/062,109
      FILING DATE: 17-MAY-1993
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/926,368
      FILING DATE: 07-AUG-1992
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: 07/242,196
      FILING DATE: 06-JUL-1988
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PCT/AU87/00401
      FILING DATE: 27-NOV-1987
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: AU P14912
      FILING DATE: 16-OCT-1987
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: AU P12570
      FILING DATE: 19-JUN-1987
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: AU PH9196
      FILING DATE: 27-NOV-1986
      ATTORNEY/AGENT INFORMATION:
      NAME: BENT, Stephen A.
      REGISTRATION NUMBER: 29,768
      REFERENCE/DOCKET NUMBER: 60042/111 BIAU
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: 202 672 5300
      TELEFAX: 202 672 5399
      TELEX: 904136
      INFORMATION FOR SEQ. ID NO: 67:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 650 amino acids
      TYPE: amino acid
      TOPOLOGY: linear
      MOLECULE TYPE: protein
      US-08-325-071-67

alignment_scores:
      Quality: 90.00      Length: 359
      Ratio: 0.612      Gaps: 22
      Percent Similarity: 40.947      Percent Identity: 21.727

alignment_block:
US-09-323-427-5/rev x US-08-325-071-67
Align seg 1/1 to: US-08-325-071-67 from: 1 to: 650

1024 TGCATTGCAAGACATGT.....TTATG 1002
      |||::: |||||
147 CysValProThrThrcysLeuArgProAspLeuThrcysLysAspLeu 163
      |||:::|||||
1001 TGAAGTCTTTATGATCAAGAGTTGCCGTATGATGTAAGTGGACGT 952
      |||:::|||||
163 sgiuysasnleuleuglnarigas.....Sera 173
      |||:::|||||
951 CAAGTTCGCGGAATTCACATTCATTCATTCATTCATTCATTCATTC 902
      |||:::|||||
173 rgcysscgslnglytrpasnthr.....AlasncysSeraIaIa 186
      |||:::|||||
901 ACATCTCTGAATCCACGTGGTATTTTGTAC..... 869
      |||:::|||||
187 ProProAlaAspser.....TyrCysSerProGlySerProLysGI 200
      |||:::|||||
868 .....ACCACTGTGTCTAT.....TTCG 850
      |||:::|||||
200 yProAspGlyGlnCysLysAsnAlaCysArgThrLysGluAlaGlyPhe 217
```

```

849 TTTCATCATTTATTTGTTACCAAGTTGATCGTATATGATGATACAA 800
      |||::: |||
217 aLcysLysHisGlyCys.ArgSerThrAspLysAlaIleTyrGluCysThr 233
      |||:::|||||
799 C.....TTTTACATGGAAGCTGTATAAACAGATAGTGCACAGATTG 759
      |||::: |||
233 sProSerGlySerThrValAlaGlnAspGlyIleThrCysLysSerIle 250
      |||::: |||
758 AGGTATCTGAAATCCACACCTGCTTTTCAACCTCAATTC.....CCGATC 712
      |||::: |||
250 eTyrThrValSerCysThrValGlnGlnLysGlnThrCysArgProThr 266
      |||::: |||
711 CCAGTATGCCGTTATGAA.....ATTGTGATGGTGCACCAACCG 671
      |||::: |||
267 GluAspCysArgValGlnLysGlyThrValLeuGlyGluCysProIrrpas 283
      |||::: |||
670 TCAACCAAGTTCAATTTGCTATCATTTGCTGACGCCAGTTTATCATAAATGGA 621
      |||::: |||
283 gln.....HisLeuValGlyAsp.....T 290
      |||::: |||
620 CATGCCATTTCTGAACCGTTGATCTTTCGCCGCGTTGCTATTCCTGC 571
      |||::: |||
290 hrcysIleSerAspCysValAspLysCys.....HisGluGlu 303
      |||::: |||
570 TTTGTGATGATGT..... 556
      |||::: |||
304 PheMetAspCysGlyValIleTyrMetAsnArgGlnSerCysTyrCysProtr 320
      |||::: |||
555 .....AACGTGATACTGTGCAAAATTTTAATGCTGATGAT 519
      |||::: |||
320 pLysSerArgLysProGlyProAsnVal.....AsnIleAsnGluC 334
      |||::: |||
518 GTGCTCTGTGATAATATTGCTAAAT.....AATTGGAA 484
      |||::: |||
334 ySLeuLeuAsnGlnTyrTyrThrValSerPheThrProAsnIleSer 350
      |||::: |||
483 TATCAACACAGATTTAAATGGCTGGCCAGAAGCTCACGTATACAAATATGC 434
      |||::: |||
351 PheAspSerAsp.....HisCysLysArgTyrGI 360
      |||::: |||
433 GGAATGCATCACACCTTTTTCATCAATGCCAGATCATGATTACCATTAAG 384
      |||::: |||
360 uAspArg.....ValLeuGlnAlaIleArgThrSerIleGlyLysGluV 375
      |||::: |||
383 AACCAATAGCAATGTTGTCGACACCAATGTTCAAGACACCAAGATTC 334
      |||::: |||
375 aPheLysValGlnIleLeu.....AsnCysThrGlnAsp..... 386
      |||::: |||
333 GGAAGCTTTAAACAGAGTGTGCGGACGACCAAAACCTGTGCACGTCCGA 284
      |||::: |||
387 .....IleLysAlaIleArgLeuIleAlaGlnLysProLeuSerLysTyrVa 401
      |||::: |||
283 ACTTCGT..... 277
      |||::: |||
401 lleuArgLysLeuGlnAlaCysGlnHisProIleGlyGlnTyrCysMetW 418
      |||::: |||
276 .....TTACTCAAGAAAAGATTCGA.....GAACG 250
      |||::: |||
418 eTyrProLysLeuLeuIleLysLysAsnSerAlaIleThrGlnIleGlnGlu 434
      |||::: |||
249 GAGAAATATCATTTGATAGCAACTGATATCAACACCCCTGAATTATGCGA 200
      |||::: |||
435 GluAsnLeuCysAspSerLeuLeuLysAsnGlnGlnAlaIleTyrLysGI 451
      |||::: |||
199 TGATTAATCAAGCTTTTCCACAGTTGAT 175
      |||::: |||
451 yGlnAsnLysCysValLysValAsp 459
      |||::: |||

seq_name: /cgn2_6/ptodata/1/laa/5A_COMB.pep:US-08-325-071-61
seq_documentation_block:
; Sequence 61, Application US/08325071
; Patent No. 5587311
```

GENERAL INFORMATION:
APPLICANT: COBON, Stewart Gary
APPLICANT: MOORE, Joanna Terry
APPLICANT: JOHNSON, Law Anthony York
APPLICANT: WILLADSEN, Peter
APPLICANT: KEMP, David Harold
APPLICANT: SRISANTHA, Alagacome
APPLICANT: RIDING, George Alfred
TITLE OF INVENTION: DNA Encoding A Cell Membrane
TITLE OF INVENTION: Glycoprotein Of A Tick Gut
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/325,071
FILING DATE: 14-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,109
FILING DATE: 17-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/926,368
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/242,196
FILING DATE: 06-JUL-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU87/00401
FILING DATE: 27-NOV-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P14912
FILING DATE: 16-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P12570
FILING DATE: 19-JUN-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PH9196
FILING DATE: 27-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 60042/111 BIAU
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 549 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-325-071-61
alignment_scores:
Quality: 89.00 Length: 359
Ratio: 0.614 Gaps: 22
Percent Similarity: 40.390 Percent Identity: 22.284
alignment_block:
US-09-323-427-5/rev x US-08-325-071-61 ..
Align seg 1/1 to: US-08-325-071-61 from: 1 to: 549

1024 TGCATTCGAAGACATGT.....TTATG 1002
|||||
66 CysValProThrThrcysLeuArgProAspLeuThrcysLysAspLeu 82
1001 TGAAGGCTCTTATGATCAAGAAGTTGCCGTATGATGATGAAGGTGAGCT 952
|||||
82 scLysAsnLeuLeuGlnArgAsp.....Sera 92
951 CAAGTGGCGGAATTCACCTCATTTGATTCATGCATGTTGGCGGTAC 902
|||||
92 rgCysCysGlnGlyTrpAsnThr.....AlaAsnCysSerAlaAla 105
901 ACAGTCTGGAATCCAGCTGATTTTGTGAACAC..... 866
|||||
106 ProProAlaAspSer.....TyrCysSerProGlySerProLysGln 119
865ACTGTTGTCA.....TTCG 850
119 yProAspGlyGlnCysIleAsnAlaCysLysMetLysGlnAlaGlyPheV 136
849 TTTTCATCATTTATTTGTTACCAAGTTGATCGTATGATCGATGATGATG 800
|||
136 alCysGlnHisGlyCys.ArgSerThrAlaLysAlaTyrGlnCysThrCys 152
799 C.....TTTACATGGAAGCTGATTAACAGCTTACGACAGATG 759
|
152 sProArgGlyPheThrValAlaGlnAspGlyIleThrCysLysSerIleS 169
758 AGGTATCTGAATCACAACCTGCTTTCAACCTCAATTGTC...CCGATG 712
:::
169 erHisThrValSerCysThrAlaGlnGlnIleGlnThrCysArgProThr 185
711 CCAGTATGCCGTTATGAA.....ATTGGATGGTGGACCAACCG 671
|||||
186 GluAspCysArgValHisLysGlyThrValLeuCysGlnCysProTrpAs 202
670 TCNACCACTTCATTTGCTATGCTATGTCAGCCAGCTTATCATTAATGGA 621
|||||
202 ngIn.....HisLeuValGlyAsp.....T 209
620 CATGCGATTCGTAACCGTTGATACTTCTGCGCGGTGTCATTCCTGC 571
|||||
209 hrcysIleSerAspCysValAspLysLysCys.....HisGlnGln 222
570 TTGTGCGATGATGCT..... 556
|||||
223 pheMetAspCysGlyValTyrMetAsnArgGlnSerCysTyrCysProTr 239
555ACGGTGAATCTGGAATTTCAATGCTGATGAT 519
239 pLysSerArgLysProGlyProAsnVal.....AsnIleAsnGlyC 253
518 GTGCTCTGATAAATAATTTGCTAAT.....AATTGGAA 484
|||
253 yslLeuLeuAsnGlnIuTyrTyrThrValSerPheThrProAsnIleSer 269
483 TATCCACAGATTTATGGCTGGCCAGAAGCTCAGCTATACAAATATGC 434
:::
270 pheAspSerAsp.....HisCysLysTrpTyrGln 279
433 GGATTCGACACGCTTTTCTATCAATGCAGATCAATCATTTACATTAAG 384
|||||
279 uAspArg.....ValLeuGlnAlaIleArgThrSerIleGlyLysGlnV 294
383 AACCAATAGCGAATGTTCGACCAATGTTGACAGACCAAGGATTC 334
:::
294 alPheLysValGlnIleLeu.....AsnCysThrGlnAsp..... 305
333 GGAAGCTTTAAACAGGTGTGCGCGACGAAACCTGCTGCGACCTCGCA 284
:::
306IleLysAlaArgLeuIleAlaGlnLysProLeuSerAsnHisVal 320

```

283 ACTTCGT..... 277
|||||
320 lleuAtgLyLeuGlnAlaLacysGlnHisProIleGlyGluTrpCysMetM 337
276 .....TTACTCAGAAAAGATCTGCA.....GAACCG 250
|||||
337 eTyrProLyLeuLeuIleLyLysAsnSerAlaThrGluIleGlu 353
249 GAGAAATATCTTGATGACGACTGATATCACAACCCCTTGAAATTAGCGA 200
|||||
354 GluAsnLeuCysAspSerLeuLeuLysAsnGlnGlnAlaIaIaTyrLysG1 370
199 TGATATATCAAGCTTTGCCAGCTTGAT 175
|||||
370 yGlnAsnLyCysValLysValAsp 378

seq_name: /cgn2.6/ptodata/1/1aa/5A_COMB.pep:US-08-325-071-63

seq_documentation_block:
; Sequence 63, Application US/08325071
; Patent No. 5587311
; GENERAL INFORMATION:
; APPLICANT: COBON, Stewart Gary
; APPLICANT: MOORE, Joanna Terry
; APPLICANT: JOHNSON, Law Anthony York
; APPLICANT: WILADSEN, Peter
; APPLICANT: KEMP, David Harold
; APPLICANT: SRISKANTHA, Alagacone
; APPLICANT: RIDING, George Alfred
; APPLICANT: RAND, Keith No. 5587311man
; TITLE OF INVENTION: DNA Encoding A Cell Membrane
; TITLE OF INVENTION: Glycoprotein Of A Tick Gut
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,071
; FILING DATE: 14-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/062,109
; FILING DATE: 17-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,368
; FILING DATE: 07-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/242,196
; FILING DATE: 06-JUL-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU87/00401
; FILING DATE: 27-NOV-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU P14912
; FILING DATE: 16-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU P12570
; FILING DATE: 19-JUN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PH9196
; FILING DATE: 27-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 60042/111 BIAU
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 650 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-325-071-63

alignment_scores:
Quality: 89.00 Length: 359
Ratio: 0.614 Gaps: 22
Percent Similarity: 40.390 Percent Identity: 22.284

alignment_block:
US-09-323-427-5/rev x US-08-325-071-63 ..
Align seg 1/1 to: US-08-325-071-63 from: 1 to: 650

1024 TGCATTCGAAGGACATGT.....TTATG 1002
|||||
147 CysValProThrThrCysLeuArgProAspLeuThrCysLysAspLeu 163
1001 TGAAGGCTCTTATGATCAAGAAGTTGCCGTAATGATGAGGTGAGCGT 952
|||||
163 sGluLysAsnLeuLeuGlnArgAsp.....SerA 173
951 CAAGTTGCCGGAATTCACCTTCATTTGATTCATGCAATGTGGCGGTAC 902
|||||
173 rGcysCysGlnGlyTrpAsnThr.....AlaAsnCysSerAlaAla 186
901 ACATCTCTGAATCCACGCGTATTTTGTAAACAC..... 866
187 ProProAlaAspSer.....TyrCysSerProGlySerProLysG1 200
865 .....ACGTGTGTCAAT.....TTCG 850
200 yProAspGlyGlnCysIleAsnAlaCysLysMetLysGluAlaGlyPheV 217
849 TTTCATTCATTTATTTGTACCAAGTTGATCGTGCATATCGACTACATG 800
|||
217 alCysGlnHisGlyCys.ArgSerThrAlaLysAlaTyrGlnCysThrCy 233
799 C.....TTTACATGGAACCTGATAAACAAGTATGACGACATTTG 759
|
233 sProArgGlyPheThrValAlaGluAspGlyIleThrCysLysSerIleS 250
758 AGGTATCTGAATACACACTGCTTTCAACACCAATTTGTC...CCGATG 712
|||
250 erHisThrValSerCysThrAlaGlnGlnLysGlnThrCysArgProThr 266
711 CCAGTATGCCGTTATGAA.....ATTTGATGTGGACCAACCGG 671
|||||
267 GluAspCysArgValHisLysGlyThrValLeuCysGlnCysProTrpAs 283
670 TCACACAGTTCAATTTGCTATATCATTTGTCACGACGTTTATCATTAATGA 621
283 nGln.....HisLeuValGlyAsp.....T 290
620 CATCGATTCCTGAACCGTTGATATCTTTCGCGGCTGTTCATTCCTGCG 571
|||||
290 hrcysIleSerAspCysValAspLysLysCys.....HisGluGlu 303
570 TTTGTGATGATGATGT..... 556
|||||
304 PheMetAspCysGlyValTyrMetAsnArgGlnSerCysTyrCysProTr 320
555 .....AACGTGATACCTGTGCAAAATTCCTAAATGCTGATGATG 519
320 pLysSerArgLysProGlyProAsnVal.....AsnIleAsnGlyC 334

```



```

651 ATGATTGGTCAGCCAGTTATATCAATAAGGACATCGGATTCGAACCGT 602
1351 .....HisLeuTrpTyrAspProGluHis 1360
601 TGATACTTCTGCGCGTGTCCATTCCTGCTTGTCCGATGAT..... 559
1360 nasrTrpGluAlaGlnIleGlnHisLeuCysThrAlaAsnSerTyrHis 1377
559 ..... 559
1377 roHisValArgLeuValGlnArgValGlnHisLeuLeuProHisVal 1393
558 ...GGTAACGGTGAATCTGTGAATTTGAATGCGATGATGCTCT 512
1394 ArgSerAsnGlyAsnProPheAspLeuAspHisAspGly...Leu 1409
511 TGATAATATTTGCTAATAT.....TGGAATATC 480
1409 uThrGluPheTyrThrAsnThrLeuSerPheGlyProAlaLeuHisTyr 1426
479 CAACAGATTTAATGCTGGCCAGAAAGCTCAGTATACAAATATCGGAT 430
1426 IaArgGluLeuValAla...GlnIleAlaHisArgTyrGlnSerMetAsp 1441
429 CGATCCAGCTTTTCTATGCATGCCAGATCAGTATTCATTAAGAAC 380
1442 IleLeuGlnIle..... 1445
379 AATATGAGATGTGTTCGACCAATGTTTCAGAACCAAGATTCGGAG 330
1446 .....GlyAlaGly 1448
329 CTGTAAACAGGTGGTGGCCAGCAAAACCTGTCAGCTGCGCAACT 280
1449 .....ThrGlyGlyAlaThrLysTyrValLeuAlaThrProGlnLeu 1462
280 ..... 280
1463 GlyPheAsnSerTyrThrTyrThrAspIleSerThrGlyPhePheGlu 1479
280 ..... 280
1479 nAlaArgGluGlnPheAlaProPheGlnAspArgMetValPheGluPro 1496
279 .....CSTTAACTCAAGAAAGATCTGACAGACCGAGAAATATC 241
1496 euAspIleArgArgSerProAlaGluGlnGlyPheGluPro...HisAla 1511
240 ATTGATGTACGAACGTATATCAACACCTTGAAATGACGATGATATCA 191
1512 TyrAspLeuIleLeuAlaSerAsnValLeuHisAlaThrProAsp.... 1526
190 AGCTTGGCCAGTGAATTAAGTCAC...CGTGCACCTTGGCAACATAATG 144
1527 ...LeuGluLysThrMetAlaHisAlaArgSerLeuLysProGlyG 1542
143 GACACCTGTAAATCTT..... 127
1542 LysGlnMetValIleLeuGluIleThrHisLysGlnHisThrArgLeuGly 1558
126 .....GTCGACGTACAAATGGAAT 107
1559 PheIlePheGlyLeuPheAlaAspTrpTrpAlaGlyValAspAspLys 1575
106 CTGCATGTACCACTTT 91
1575 gCysThrGluProPhe 1580
seq_name: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:US-08-637-640-2
seq_documentation_block:
; Sequence 2, Application US/08637640
; Patent No. 5849541

```

```

; GENERAL INFORMATION:
; APPLICANT: VINCI, VICTOR A.
; APPLICANT: CONDER, MICHAEL J.
; APPLICANT: MCADA, PHYLIS C.
; APPLICANT: REEVES, CHRISTOPHER D.
; APPLICANT: DAVIS, CHARLES R.
; APPLICANT: HENDRICKSON, LEE E.
; APPLICANT: RAMBOSEK, JOHN
; TITLE OF INVENTION: DNA ENCODING TRIOL POLYKETIDE SYNTHASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHRISTINE E. CARTY
; STREET: P.O. BOX 2000; 126 E. LINCOLN AVENUE
; CITY: RAHWAY
; STATE: NJ
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,640
; FILING DATE: 23-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/148,132
; FILING DATE: 01-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: CARTY, CHRISTINE E.
; REGISTRATION NUMBER: 36,099
; REFERENCE/DOCKET NUMBER: 19076
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-594-6734
; TELEFAX: 908-594-4720
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3038 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: TPKS Protein
; US-08-637-640-2

alignment_scores:
    Quality: 87.00      Length: 472
    Ratio: 0.481        Gaps: 24
    Percent Similarity: 38.347    Percent Identity: 19.703

alignment_block:
US-09-323-427-5/rev x US-08-637-640-2 ..
Align seg 1/1 to: US-08-637-640-2 from: 1 to: 3038

1146 ATTGCTTTGTACTACTTATGTCATTTGCTTATTCGATCCGGTTGA 1097
1175 ValAlaPheGlnThrValIleGlyAla.....TyrSerSerPro.... 1187
1096 CAATGCTGTCAAGGTGACGCAAGAAATGGAATGT.....GGACCACTT 1053
1188 .....GlyAspArgArgLeuArgCysLeuTyrValProHis 1200
1052 CAATA..... 1048
1200 IaValAspArgIleThrLeuValProSerLeuCysLeuAlaThrAlaGlu 1216
1047 .....ACATTCATTTTAATACAGCTAATGCAATTCGAAGGACA 1010

```

```

1217 SerGlyCysGluValAlaPheAsnThrIleAsnThrTyrAspLysG1 1233
1009 TGTATTGTAAGAGT.....CTTTATGATCAGAAAGTGGCCGTA 969
      |||:::||||
1233 yAspTyrLeuSerGlyAspIleValAlaPheAspAlaGlu..... 1246
      |||:::||||
968 ATGATGAAGGTGACGTCAGATTGCCGGAATTCACATT...CCATTGAT 922
      |||:::|||||
1247 ..GlnThrThrLeuPheGlnValGlnAsnIleThrPheLysProPheSer 1262
      |||:::|||||
921 TCATGCATTTGCGCGTACACGATCTCTGAATCCAGTGGTATTTT.. 874
      |||:::|||||
1263 ProProAspIleSerThrAspHisAlaMetPheAlaArgTyrSerTyrG1 1279
      |||:::|||||
873 ....GTAACAACAACGTGTGTCATTGTCATTCCATTTATTTGTTACA 828
      |||:::|||||
1279 yProLeuThrProAspSerLeuAspAsnProGluTyrTyrPheAlaThr 1296
      |||:::|||||
827 AAGTTGATCGT.....GCATATGACAGTACATGCTTTTAC 793
      |||:::|||||
1296 IacIleAspLysGluAlaIleProIleIleGluArgIleValTyrPheTyr 1312
      |||:::|||||
792 ATGGAACGCTATAAACAGTTAGTCACACAGATGAGTATCTGAATCAC 743
      |||:::|||||
1313 Ile.....ArgSerPheLeuSerGlnLeuThrLeuGlnGluArgG1 1326
      |||:::|||||
742 AACTGCT.....TTCAAACTCAATTGTCGCGATGCCAGTATGCC 702
      |||:::|||||
1326 ngIleAlaIlePheHisLeuGlnLysGlnIleGluTyrPheGlu..... 1340
      |||:::|||||
701 GTTATGAATTTTGGATGTGTGACCAACCGTCACACGATTCATTTGCT 652
      |||:::|||||
1341 ....GlnValLeuAlaSerAlaLysGlnLysArg..... 1350
      |||:::|||||
651 ATCATGTGTACGCCAGTTTATCTAAATGACATGCCATTCGTAACCGT 602
      |||:::|||||
1351 .....HisLeuTyrTyrAspProLysTyrPheLys 1360
      |||:::|||||
601 TGATTACTTCTGCGCGTGTCCATTCCTGCTTTCGATGAT..... 559
      |||:::|||||
1360 nAspThrGlnAlaGlnIleGlnHisLeuGlnLysThrAlaAsnSerTyrHisP 1377
      |||:::|||||
559 ..... 559
      |||:::|||||
1377 roHisValArgLeuValGlnArgValGlyGlnHisLeuLeuProThrVal 1393
      |||:::|||||
558 ..GGTAACGCTGATACTGTGGAATTCATAATGCTGATGATGCTCT 512
      |||:::|||||
1394 ArgSerAsnGlyAsnProPheAspLeuLeuAspHisAspLys...LeuLe 1409
      |||:::|||||
511 TGATTAATATTTCCTAAATAT.....TTGGAATATC 480
      |||:::|||||
1409 uThrGluPheTyrThrAsnThrLeuSerPheGlyProAlaLeuHisTyrA 1426
      |||:::|||||
479 CAACAGATTTAATGCTGCGCAAGAACTCAGTATACAAATATGCGGAT 430
      |||:::|||||
1426 IaArgGluLeuValAla...GlnIleAlaHisArgTyrGlnSerMetAsp 1441
      |||:::|||||
429 CGATCAGAGCTTTCTATCAATGCCAGATCATTAACCTTAAGAAC 380
      |||:::|||||
1442 IleLeuGluIle..... 1445
      |||:::|||||
379 AATATGCGAATGTGTCGACACAAATGTCAGAACACAGATTCGAG 330
      |||:::|||||
1446 .....GlyAlaGly.. 1448
      |||:::|||||
329 CTGTTAAACAGGTGTGCGCGACGAAACCTGCTCAGCTGCGCAACTT 280
      |||:::|||||
1449 .....ThrGlyGlyAlaThrLysTyrValLeuAlaThrProGlnLeu 1462
      |||:::|||||
280 ..... 280
      |||:::|||||
1463 GlyPheAsnSerTyrThrTyrThrAspIleSerThrGlyPhePheGlnG1 1479

```

```

280 ..... 280
1479 nAlaArgGluGlnPheAlaProPheGluAspArgMetValPheGluProL 1496
      |||:::|||||
279 .....CGTTACTCAAGAAAAGATCTGCAGAACCGAGAAATATC 241
      |||:::|||||
1496 euAspIleArgArgSerProAlaGluGlnGlyPheGluPro...HisAla 1511
      |||:::|||||
240 ATTGATGTGAACCTGATATCAACACCCCTGAAATTAAGCATATATCA 191
      |||:::|||||
1512 TyrAspLeuValIleLeuAlaSerAsnValLeuHisAlaThrProAsp.... 1526
      |||:::|||||
190 AGCTTGGCCAGTATTTACCTCAC...CGTGACCTTCTGCACATATATG 144
      |||:::|||||
1527 ....LeuGlnLysThrMetAlaHisAlaArgSerLeuLeuLysProGlyG 1542
      |||:::|||||
143 GACAACTGTATATCTT.....GCTGACGTACAAATGGAAT 107
      |||:::|||||
1542 LysGlnMetValIleLeuGluIleThrHisLysGlnHisThrArgLeuGly 1558
      |||:::|||||
126 .....GCTGACGTACAAATGGAAT 107
      |||:::|||||
1559 PheIlePheGlyLeuPheAlaAspTyrTyrPheAlaGlyValAspAspLys 1575
      |||:::|||||
1575 gCysThrGluProPhe 1580

```

```

seq.name: /cgn2-6/plodata/1/aa/5B_COMP.pep:US-08-540-804-12

```

```

seq.documentation_block:

```

```

: Sequence 12, Application US/08540804
: Patent No. 5919666

```

```

: GENERAL INFORMATION:

```

```

: APPLICANT: Young, Richard A.

```

```

: APPLICANT: Koleske, Anthony J.

```

```

: APPLICANT: Thompson, Craig M.

```

```

: APPLICANT: Chao, David M.

```

```

: TITLE OF INVENTION: No. 5919666el Factors Which Modify Gene

```

```

: NUMBER OF SEQUENCES: 39

```

```

: CORRESPONDENCE ADDRESS:

```

```

: ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.

```

```

: STREET: Two Millitia Drive

```

```

: CITY: Lexington

```

```

: STATE: Massachusetts

```

```

: COUNTRY: USA

```

```

: ZIP: 02173

```

```

: COMPUTER READABLE FORM:

```

```

: MEDIUM TYPE: Floppy disk

```

```

: COMPUTER: IBM PC compatible

```

```

: OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

: SOFTWARE: Patent Release #1.0, Version #1.30

```

```

: CURRENT APPLICATION DATA:

```

```

: APPLICATION NUMBER: US/08/540,804

```

```

: FILING DATE: 11-OCT-1995

```

```

: CLASSIFICATION: 424

```

```

: PRIOR APPLICATION DATA:

```

```

: APPLICATION NUMBER: US 08/521,872

```

```

: FILING DATE: 21-AUG-1995

```

```

: PRIOR APPLICATION DATA:

```

```

: APPLICATION NUMBER: US 08/218,265

```

```

: FILING DATE: 25-MAR-1994

```

```

: ATTORNEY/AGENT INFORMATION:

```

```

: NAME: Granahan, Patricia

```

```

: REGISTRATION NUMBER: 32,227

```

```

: REFERENCE/DOCKET NUMBER: WH194-0342

```

```

: TELECOMMUNICATION INFORMATION:

```

```

: TELEPHONE: 617-861-6240

```

```

: TELEFAX: 617-861-9540

```

```

: INFORMATION FOR SEQ ID NO: 12:

```

```

: SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 1226 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-540-804-12

```

```

alignment_scores:
  Quality: 85.50      Length: 390
  Ratio: 0.457        Gaps: 13
Percent Similarity: 47.949   Percent Identity: 19.744

```

```
alignment_block:
```

```
US-09-323-427-5/rev x US-08-540-804-12
```

```
Align seg 1/1 to: US-08-540-804-12 from: 1 to: 1226
```

```

1148 TTTATGCTTCTGCTACTACCTATTGCTATTGCTTATTCGATTCGCGTT 1099
      ::::: |||::: ::::: |||::: |||::: |||:::
358 MetMetProSerLeuTyrArgLeuValLeuLeuThrTyrGlyI 374
      ::::: |||::: ::::: |||::: |||::: |||:::
1098 GACAAATGCTGCAAGTGCAGCAAAATGATGAGCAACTTCAT 1049
      ::::: |||::: ::::: |||::: |||::: |||:::
374 eilelYsValProThrTyrIleArgLysLeuIle..... 385
1048 AACATCAATTTTAAATACAGTAATGCAATGCAAGCATGTTATGTA 999
      ::::: |||::: ::::: |||::: |||::: |||:::
386 ..SerSerGlyLeuLeuTyrLeuGlnAspSerAsnSplysPheValHis 401
998 AAGGCTTTATGATCAAG..AAGGTTGCCGTAATGATGAAGTGACGT 952
      ::::: |||::: |||::: |||::: |||::: |||:::
402 ValGlnLeuLeuLeuLeuLeuLysIleSerProLeuMetLys..... 415
951 CAAGTGGCCGGAATTCACCTTCATTGATTCATGCAATGTTGCCGTAC 902
      ::::: |||::: ::::: |||::: |||::: |||:::
416 .....SerGlnTyrAsnMetValIleuArgAsnValMetGluT 428
901 ACGATCTCTGAATCCACGTCGATTTTGTAAACAACATG..... 861
      ::::: |||::: |||::: |||::: |||::: |||:::
428 YrsPValLysPheTyrGlnIlePheAsnPhaSpGlnLeuValGluIle 444
860 .....TTCATTCCTTCGTTTCATTCATTCATTTT 835
445 ThrGlnGlnIleLysMetArgIleLeuSerAsnAspIleThrAsnLeuG 461
834 GTTACCAAGTGTGATCGTCATATCGATCAATGCTTTACATGCAAGC 785
      ::::: |||::: |||::: |||::: |||::: |||:::
461 nLeuSerLysThrProLeuSerIleLysIleMetValAlaGluTyrIle 478
784 TG.....ATTAACAGTTAGTGCACAGATTGAGGTAT 753
      ::::: |||::: |||::: |||::: |||::: |||:::
478 euserHisLeuCySerGlyIleLeuSerSerValAsnArgThrValLeu 494
752 CTGAATACACACTGCTTTC.....AAACTCAATTTGCCGATGCCA 709
      ::::: |||::: |||::: |||::: |||::: |||:::
495 LeuLysIlePheLysIlePheCysIleAspLeuGlnValPheHisSph 511
708 GTATGCCGTTATGAATTTGATGATGAGCAACCGTCACACAGTTCA 659
      ::::: |||::: |||::: |||::: |||::: |||:::
511 ePheLysThrIleGlnPheIleValIleValIleGlnLeuLeuSerAspIleG 528
658 ATTTGCTA.....TCATTGGTC 642
      ::::: |||::: |||::: |||::: |||::: |||:::
528 lserLeuGlnValAlaLeuMetAspIleLeuLeuCySTyrGlnLysLeuPhe 544
641 ACCCACTTTATCATTAATGACATGCGATTCG...AACCGTTGATA.. 597
      ::::: |||::: |||::: |||::: |||::: |||:::
545 SerGlnPhe...IleAsnAspHisIleLeuPheThrLysThrPheIlePhe 560
596 .....CTTTCTGCGGGGTTGCCATTCCTCGTTTGCGCATGA 560
      ::::: |||::: |||::: |||::: |||::: |||:::
560 eileTyrLysLysValLeuLysGlnLysAspValProAlaTyrAsnValT 577

```

```

559 TGCTAACGGTACTGCTGGAATTTCTAAATGCTGATGATGCTCTTG 510
      ::::: |||::: |||::: |||::: |||:::
577 hrserPheMetProPheTyrPlySphPheMetLysAsnPhaProPheVal 593
      ::::: |||::: |||::: |||::: |||:::
509 ATAAATATTTTGCTAAATATTGGAATATCCACACAGATTAAAGCGCTGC 460
      ::::: |||::: |||::: |||::: |||:::
594 LeuLysValAspAsnAspLeuArgIleGlnLeuGlnSerValTyrAsnAs 610
459 CAAGAAGCTCAGCTAATACAAATATGCGATGATGCACAGCTTTTCATCA 410
      ::::: |||::: |||::: |||::: |||:::
610 pGluLysLeuLysThrGlnLysLeuLysAsnAspLysSerGlnValLeu 627
409 ATGCCAGATCAGTATTACCATTAAGAACA..... 378
      ::::: |||::: |||::: |||::: |||:::
627 yValTyrSerMetIleAsnAsnSerAsnGlnAlaValGlyGlnThrTyr 643
377 ..ATAGCGAATGTTCCGACCAATGTT..... 351
      ::::: |||::: |||::: |||::: |||:::
644 AsnPhaProGlnValPheGlnValAsnIleArgPheLeuLeuHisAsnSe 660
350 .....CAGAACCAACAGGATT 335
      ::::: |||::: |||::: |||::: |||:::
660 rGluIleIleAspThrAsnThrSerLysGlnPheGlnLysAlaArgAsnA 677
334 CGGAGCTGTTAAACAGGTCGCGCCGACGCAAAACCTGTCAGCTGCC 285
      ::::: |||::: |||::: |||::: |||:::
677 snValMetLeuLeuIleAlaThrAsnLeuLysGlnLysPheMet 693
284 AACTGCTTACTCAGAAAGATCTGCAGAACCGAGATA..... 243
      ::::: |||::: |||::: |||::: |||:::
694 SerIlePheLeuLysArgLysAspPheThrAsnLysAsnLeuIleGlnLe 710
242 .....TCATTGATGACAACTGATATACACACCTTGAAT 206
      ::::: |||::: |||::: |||::: |||:::
710 uIleSerLeuLysLeuLeuThrPheGlnValThrGlnAsnValLeuGly 727
205 TACGATGATATATCAAGCTT 186
      ::::: |||::: |||::: |||::: |||:::
727 eGluTyrIleIleIleArgLeu 733

```

```
seq_name: /cgn2_6/plodata/1/laa/5B_COMB.pep:US-08-218-265-12
```

```

seq_documentation_block:
; Sequence 12, Application US/08218265
; Patent No. 5922585
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A.
; APPLICANT: Thompson, Craig M.
; TITLE OF INVENTION: No. 5922585el Factors Which Modify Gene
; TITLE OF INVENTION: Transcription and Methods of Use Thereof
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millita Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: US
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/218, 265
; FILING DATE: 25-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI94-03
; TELECOMMUNICATION INFORMATION:

```



```
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPND-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-604A-6

alignment_scores:
Quality: 85.00 Length: 196
Ratio: 0.825 Gaps: 9
Percent Similarity: 52.551 Percent Identity: 22.449

alignment_block:
US-09-323-427-5/rev x US-08-480-604A-6 ..
Align seg 1/1 to: US-08-480-604A-6 from: 1 to: 2710

633 TATCATAAATGACATGCGATCTGAAACCGTGGATCTTTCGGCGGT 584
|||||
188 TyTtYlYsSerGlnIleasnLysProThValProThIleasnspil 204
|||||
583 TGTSCATTCCTGCTTGTGCGATGATGTAACGGTGATCTGTGGAATTC 534
|||||
204 etIeLysSerHIsleuValSerGu.....T 213
533 TAAATGCTATGATGTGCTCTGATTAATATTGCTAAATATTGGAA 484
|||||
213 yTasnArGspRglurHValIleuGlnSerTyArGrThrAsnSerLeuArG 229
|||||
483 TATCCAACAGATTATGCTGGCCCAAGAGCTCAGTATACAATATGC 434
|||||
230 .....Lys..IleA 232
433 GATATCGATCACAGCTTTCTATCAATGCCAGATCAGTATACCATTAAG 384
|||||
232 snSerAsnHIsGlyIleAspIleArgAlaIsnSerLeuPhe...ThnGlu 247
383 AACCAAAATAGCAATGTGTGACCAACAATGTTCCAGAACCAAGATTTC 334
|||||
248 GlnGlnIleuLeuAsnIleTySerGlnIleuLeuAsnArgGlyAsnIle 264
333 GGAGCTGTTAAACAGGTGTGCGCCGACGAAACCTGCTCAGCTGCGCA 284
|||||
264 uAlaAlaAlaIleSerAspIleVal.....ArgLeuLeuAlaIleuLysA 278
```

```
283 ACTTC.....GTTTACTCAAGAAAGATCTGCAGAACCGGAGAAATATCA 240
|||||
278 snPheGlyGlyValTyLeuAspValasPmetLeuProGlyIleHIsSer 294
|||||
239 TTGATGTACGAACTGATATCAACACCC..... 213
|||||
295 AspLeuPheLysThrIleSerArgProSerSerIleGlyLeuAspArgTr 311
|||||
212 .....TGAAATTAAGCATGATTAATCAAGCTTGGCAGTTGAT..... 174
311 pGluMetIleuLysLeuGlnAlaIleMetLysTyLysLysTyIleAsnA 328
173 ..TACGTACCGCTGCACCTC.....TGCAACATTAATGACACAC 138
|||||
328 snTyThrSerGluAsnPheAspLysLeuAspGlnIleuLysAspAsn 344
137 CTGTAATCTTGCTGCGAGTACAAATGCAATCTGCA 102
345 PheLysLeuIleIleGlnSerLysSerGluLysSer 356

seq_name: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:US-08-405-496A-6

seq_documentation_block:
Sequence 6, Application US/08405496A
Patent No. 5919665
GENERAL INFORMATION:
APPLICANT: WILLIAMS, JAMES A.
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSER: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPND-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-405-496A-6
```

alignment_scores:
 Quality: 85.00 Length: 196
 Ratio: 0.825 Gaps: 9
 Percent Similarity: 52.551 Percent Identity: 22.449

alignment_block:

US-09-323-427-5/rev x US-08-405-496A-6 ..

Align seg 1/1 to: US-08-405-496A-6 from: 1 to: 2710

```

633 TATCATTAATGACATGCGATTCTGAAACCGTTGATTAATTTCTGCGCGT 584
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
188 TyrTyrLysSerGlnIleAsnLysProThrValProThrIleAspAspI1 204
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
583 TGTCCATTCCTGCTTGTGGATGATGTAAGCGTGATCTGTGGAATTC 534
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
204 eLlEysSerHisLeuValSerGlu.....T 213

533 TAAATGCTGATGATGCTGCTTGTGATAATATTGCTAAATATTGGA 484
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
213 yRAsnArgAspGlnThrValLeuGlnUserTyrArgThrAsnSerLeuArg 229
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
483 TATCCACACAGATTTAATGCTGGCCAGACAGCTACGATACAAATATGC 434
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
230 .....Lys.IleA 232

433 GGATGCGATCAGACGCTTTCTATCATGCGACAGTACGATTAACATTAAAG 384
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
232 snSerAsnHisGlyIleAspIleArgAlaAsnSerLeuPhe...Thriglu 247
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
383 AACCAATAGCGAATGTGTGCGACCAATGTTGACAAACACACAGAGATTTC 334
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
248 GlnGlnLeuLeuAsnIleTyrSerGlnGlnLeuLeuAsnArgGlyAsnLe 264
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
333 GGAGCTGTTAAACAGGTGTCGCCAGCAAAACCTGCTGCGAGTCGCGCA 284
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
264 uAlaIleAlaSerAspIleVal.....ArgLeuLeuAlaLeuLysA 278
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
283 ACTTC.....GTTACTCAAGAAAGATCTGCAGACCGGAGATATCA 240
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
278 snPheGlyGlyValTyrLeuAspValAspMetLeuProGlyIleHisSer 294
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
239 TTGATGATGAGCAATGATATCAACACCC..... 213
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
295 AspLeuPheLysThrIleSerArgProSerSerIleGlyLeuAspArgTr 311
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
212 .....TTGAATTAGCGATGATATCAAGCTTTGCCAGTTGATT.... 174
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
311 pGlnMetIleLysLeuGlnAlaIleMetLysTyrLysTyrIleAsnA 328
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
173 ..TAGCTACCGTGCACTTC.....TGCAACATATGACACAC 138
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328 snTyrThrSerGlnAsnPheAspLysLeuAspGlnGlnLeuLysAspAsn 344
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
137 CTGTAATCTTCTGCTGACGTACAAATGCAATGCA 102
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
345 PheLysLeuIleIleGlnSerLysSerGlnLysSer 356
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

seq.name: /cgn2_6/ptodata/1/laa/5A_COMB.pep:US-08-325-071-65

seq_documentation_block:

; Sequence 65, Application US/08325071
 ; Patent No. 5587311

; GENERAL INFORMATION:

; APPLICANT: COBON, Stewart Gary
 ; APPLICANT: MOORE, Joanna Terry
 ; APPLICANT: JOHNSON, Law Anthony York
 ; APPLICANT: WILLADSEN, Peter
 ; APPLICANT: KEMP, David Harold
 ; APPLICANT: SRISKANTHA, Alagaccone
 ; APPLICANT: RIDING, George Alfred
 ; APPLICANT: RAND, Keith No. 5587311man

;; TITLE OF INVENTION: DNA Encoding A Cell Membrane
 ;; TITLE OF INVENTION: Glycoprotein Of A Tick Gut
 ;; NUMBER OF SEQUENCES: 71
 ;; CORRESPONDENCE ADDRESS:
 ;; ADDRESSEE: Foley & Lardner
 ;; STREET: 3000 K Street, N.W.
 ;; CITY: Washington, D.C.
 ;; COUNTRY: USA
 ;; ZIP: 20007-5109

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; OPERATING SYSTEM: IBM PC compatible

;; SOFTWARE: Patent In Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/325,071

;; FILING DATE: 14-OCT-1993

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 08/062,109

;; FILING DATE: 17-MAY-1993

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/926,368

;; FILING DATE: 07-AUG-1992

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: 07/242,196

;; FILING DATE: 06-JUL-1988

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: PCT/AU87/00401

;; FILING DATE: 27-NOV-1987

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: AU P14912

;; FILING DATE: 16-OCT-1987

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: AU P12570

;; FILING DATE: 19-JUN-1987

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: AU PH9196

;; FILING DATE: 27-NOV-1986

;; ATTORNEY/AGENT INFORMATION:

;; NAME: BENT, Stephen A.

;; REGISTRATION NUMBER: 29,768

;; REFERENCE/DOCKET NUMBER: 60042/111 BIAU

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: 202 672 5300

;; TELEFAX: 202 672 5399

;; TELEX: 904136

;; INFORMATION FOR SEQ ID NO: 65:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 620 amino acids

;; TYPE: amino acid

;; TOPOLOGY: linear

;; MOLECULE TYPE: protein

;; US-08-325-071-65

alignment_scores:

Quality: 84.00 Length: 359

Ratio: 0.575 Gaps: 22

Percent Similarity: 40.669 Percent Identity: 21.448

alignment_block:

US-09-323-427-5/rev x US-08-325-071-65 ..

Align seg 1/1 to: US-08-325-071-65 from: 1 to: 620

```

1024 TGCATTCGACGACATGT.....TTATG 1002
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
117 CysValProThrThrCysLeuArgProAspLeuThrCysLysAspLeu 133
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1001 TGAAGGCTTTATGATCAAGAGGTGCCGATATGATGAAGGTGGACGT 952
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
133 sGlnLysAsnLeuLeuGlnArgAsp.....Sera 143

```

951 CAAGTCCCGGAATTTCACCTTCATTGATTCATGCAATGTTGCCGTAC 902
|||:::|||||
143 rGyScyGlnGlnGlyTrpAsnThr.....AlaAsnCysLeuAlaIa 156
901 ACGATCTCGAATCCACGTGGTATTTTGTAC..... 869
157 ProProAlaAspSer.....TyrCysSerProGlySerProLysG1 170
868AACAACTGTTGTCAT.....TTGC 850
170 yProAsnArgLysGlnCysLysAsnAlaCysArgThrLysGlnAlaGlyPheV 187
849 TTTTCATCATATTGTTTACCAAGTGTATGCGCATATCGAGTCAATG 800
||:::|||||
187 a1CysAluHisGlyCys..ArgSerThrAspLysAlaTyrGlnCysThrCy 203
799 C.....TTTACATGGAGAGCTGATAAACAGTTAGTCACAGATG 759
|:::|||||
203 sProSerGlySerThrValAlaGlnAspGlyLeuThrCysLysSerIles 220
758 AGGTATCTGAATTCACAACTGCTTTCAAATCAATATGTC...CCGATG 712
:::|||||
220 eTtyrThrValSerCysThrValGlnGlnLysGlnThrCysArgProThr 236
711 CCAGTATGCCGTATGAA.....ATTGATGCGTGACCAACGG 671
|||||:::|||||
237 GluAspCysArgValaGlnLysGlyThrValLeuCysGlnCysProThrAs 253
253 nGln.....HisLeuValGlyAsp.....L 260
620 CATGGATCTCGAAACCGTTGATCTTTCGCCGGTGTGTCATCTCTGC 571
|||:::|||||
260 yScySl1eSerAspCysValAspLysLysCys.....HisGlnGlu 273
570 TTTGTGATGATGCT..... 556
274 PheMetAspCysGlyValIyrMetAsnArgGlnSerCysTyrCysProTr 290
555AACGGTATCTGTGGAATTCGAATGCTGATGAT 519
290 pLysSerArgLysProGlyProAsnVal.....AsnIleAsnGlnC 304
518 GTGCTCTTGATAAATATTGCTAAAT.....AATTTGAA 484
||:::|||||
304 ySeLeuLeuAsnGlnIuTyrTyrThrValSerPheThrProAsnIleSer 320
483 TATCCAAACGATTTAATGGCTGGCCAGAACGTCACGATACAAATATGC 434
:::|||||
321 PheAspSerAsp.....HisCysLysArgTyrG1 330
433 GGATGATCACAGCTTTTCTATCAATGCCAGATCAGATTACCATTAAG 384
|||||:::|||||
330 uAspPheG.....ValLeuGlnAlaIleArgThrSerIleGlyLysGlnV 345
383 AACCAATAGCGAATGTGTTCCAGCCACAATGTTCAAGACACAGGATTC 334
:::|||||
345 a1PheLysValaGlnIleLeu.....AsnCysThrGlnAsp..... 356
333 GGAAGCTGTTAAACAGGTGTGCGCAGCAAAACCTGCTGCAGCTCGCA 284
:::|||||
357IleLysAlaArgLeuIleAlaGlnLysProLeuSerLysTyrVa 371
283 ACTTGGT..... 277
|||
371 lleuArgLysLeuGlnAlaCysGlnHisProIleGlyGluTrpCysMetM 388
276TTACTCAAGAAAAGATCTGA.....GAACCG 250
||:|||||
388 eTtyrProLysLeuLeuIleLysLysAsnSerIleArgIuIleGlnGlu 404
249 GAGAAATATCATTTGATGTACGAACGTGATATCAACACCCTTGAATTAGCGA 200

|||||::: |||
405 GluAsnLeuCysAspSerLeuLeuLysAsnGlnGlnAlaAlaTyrLysG1 421
199 TGATAATCAAGCTTTGCCAGTTGAT 175
:::|||||
421 yGlnAsnLysCysValLysValAsp 429

THIS PAGE BLANK (USPTO)


```

640 GCCAGTTTATCATTAATGACATGCGATTCTGAAACCGTTGATTAATCTTCT 591
1   |||||||||||||||||||||||||||||||||||||||||||||||||||
174 nglvaltlytnhlsystrphrcysaspsersglutlntrhAspThrPheC 191
590 GCGCGGTTGTCATTCCTGCTTTGTCGATGATGTTAACGGTGTACTGTCG 541
   |||||||||||||||||||||||||||||||||||||||||||||||||||
191 ysAlaValAlaHisSerCysThrValAspAspGlyAsnGlyAspThrVal 207
540 GAAATCTTAATGCGTATGATGATGCGTCTTGAATTAATATTTGCTAATA 491
   |||||||||||||||||||||||||||||||||||||||||||||||||||
208 GlnIleuAsnGlnGluGlyCysAlaLeuAspLysPheLeuAsnAs 224
490 TTGGAATATCCAAAGATTTAATGCGTGGCCAAAGATCGATATACA 441
   |||||||||||||||||||||||||||||||||||||||||||||||||||
224 nleuglutyrrProThrAspIleuMetAlaGlnGlnGlnAlaHisValTyl 241
440 AAATATGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 391
   |||||||||||||||||||||||||||||||||||||||||||||||||||
241 ystYrAlaAspArgSerGlnLeuPheTyrGlnCysGlnIleSerIleThr 257
390 ATTAAGAACCAATATGCGAATGTGTCGACACAAATGTTTACAGAACACA 341
   |||||||||||||||||||||||||||||||||||||||||||||||||||
258 IleLysAspProGlySerGlnCysAlaArgProThrCysSerGlnProG 274
340 AGGATTCGAGAGCTGTTAAACAGAGTGGTGGCCGACGACAAACSTGTCAG 291
   |||||||||||||||||||||||||||||||||||||||||||||||||||
274 nglrPheGlyAlaValLysGlnAlaGlyAlaGlyAlaHisAlaAla 291
290 CT..... 289
291 laAlaProGlnAlaGlyAlaGlnGluValGlnAlaAlaProValGlyAla 307
289 ..... 289
308 AlaProValAlaAlaProValAlaAlaAlaAlaAlaAlaProAlaValPr 324
288 ..... GCGCACTTCGTTTACTC... AAGAAAGATCT... G 258
324 oATgAlaThrIleuAlaGlnLeuArgLeuLeuArgLysArgSerPheG 341
257 CAGAACCGGAGATATCATTTGATGATGACGAACGTATATCAACCCCTGAA 208
   |||||||||||||||||||||||||||||||||||||||||||||||||||
341 LysGlnAsnGlnGlyIleLeuAspValArgValGlnIleAsnThrLeuAsp 357
207 ATTAGCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 158
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
358 IleMetGlnGly..... AlaSerProSerAlaProGlnAlaAlaAla 372
157 TCTGCACATATGACAACTGTAATACTTCTGTCGACGATCAAAATGGAA 108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
372 uValSerGln..... GluSerValArgArgAlaThrSerThrGlyI 387
107 TCTGCATGTCACCATTTGGCTCTCAATGTTATGGTTTAAGCATTTGCA 58
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
387 leSerSerThrProIleGlyLeuProSerPheLeuGlnMetArgThrIle 403
57 TTGATTGCTGCGCTCATTTATACCATTTGCTTAAATTTGTCGCA 13
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
404 ValAlaThrAlaLeuSerAlaThrIlePheTyrValAlaAlaArgPro 418

```

seq_name: p1r2:S27799

seq_documentation_block:
 cuticlin 1 - Caenorhabditis elegans (fragment)
 C:Species: Caenorhabditis elegans
 C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Sep-1997
 C:Accession: S27799
 R:Sebastiano, M.; Lassandro, F.; Bazicalupo, P.
 submitted to the EMBL Data Library, July 1991
 A:Description: Cuf-1 a Caenorhabditis elegans gene coding for a dauer specific non colla
 A:Reference number: S27799
 A:Accession: S27799
 A:Molecule type: DNA

```

A:Residues: 1-308 <SEB>
A:Cross-references: EMBL:M55997; NID:g156271; PID:g156272
C:Genetics:
A:Gene: Cuf-1
A:Introns: 245/3

alignment_scores:
  Quality: 897.50      Length: 302
  Ratio: 3.989         Gaps: 5
  Percent Similarity: 74.503   Percent Identity: 60.596

alignment_block:
  US-09-323-427-5/rev x S27799 ..
  Align seg 1/1 to: S27799 from: 1 to: 308

801 TCGTTTATCATGGAACGTGTAACAGATTAGTGCACAGATTGAGTATC 752
   |||||||||||||||||||||||||||||||||||||||||||||||||||
6   CysPheTyrMetGlnSerAspLysThrValSerThrGlnIleGluValSe 22
751 TGAATTCACACTGCTTTCAACTCAATATGTCGATGCGATGCGATGCGC 702
   |||||||||||||||||||||||||||||||||||||||||||||||||||
22   rAspLeuThrThrAlaPheGlnThrGlnValAlaProMetProValCysL 39
701 GTTATGAATTTTGGATGGTGCACAAACCGTCAACAGTTCAATTGCT 652
   |||||||||||||||||||||||||||||||||||||||||||||||||||
39   ystYrGlnIleuAspArgGlyLysProSerGlnProIleGlnPheAla 55
651 ATCATTTGTCAGCCAGTTTATCATTAATGACACATGCGATTTCAACCG 602
   |||||||||||||||||||||||||||||||||||||||||||||||||||
56   ThrIleGlnGlnGlnValAlaTyrHisLysTrpThrCysAspSerGlnThr 72
601 TGATACTTCTGCGCGTTCCTCATTCCTGCTTGTGCGATGATGATGATG 552
   |||||||||||||||||||||||||||||||||||||||||||||||||||
72   rAspThrPheCysAlaValAlaHisSerCysThrValAspAspGlyAsnG 89
551 GTGATCTGTGGAATTTGTAATGCTGATGATGATGCTCTTGAATATAT 502
   |||||||||||||||||||||||||||||||||||||||||||||||||||
89   LysPheThrValGlnIleLeuAsnGlnGlnGlyCysAlaLeuAspLysPhe 105
501 TTGCTAATATATTGGATATTCACACAGATTTAATGCTGGCCCAAGAGC 452
   |||||||||||||||||||||||||||||||||||||||||||||||||||
106   LeuLeuAsnAsnLeuGlnTyrProThrAspLeuMetAlaGlnGlnAla 122
451 TCACGTATACAAATATGCGGATGATGATGATGATGATGATGATGATG 402
   |||||||||||||||||||||||||||||||||||||||||||||||||||
122   ahIsValTyrLysTyrAlaAspArgSerGlnLeuPheTyrGlnCysGlnI 139
401 TCAGTATTACCATTTAAAGAACCAATAGCGAATGTGTCGACCAACATGT 352
   |||||||||||||||||||||||||||||||||||||||||||||||||||
139   leSerIleThrIleLysAspProGlySerGlnCysAlaArgProThrCys 155
351 TCAGAACCCAAAGATTTGGAGCTGTTAAACAGACGTTTCCATCAATGCA 302
   |||||||||||||||||||||||||||||||||||||||||||||||||||
156   SerGlnProGlnGlyPheGlyAlaValLysGlnAlaGlyAlaGlyAla 172
301 ACCTGCTGCAGCT..... 289
   |||||||||||||||||||||||||||||||||||||||||||||||||||
172   ahIsAlaAlaAlaAlaAlaProGlnAlaGlyValGlnGluValAlaAla 189
289 ..... 289
189   roValGlyAlaAlaProValAlaAlaProValAlaAlaAlaAlaAla 205
288 ..... GCGCACTTCGTTTACTC... AAGAA 266
206   ProAlaValProArgAlaThrIleuAlaGlnGlnLeuArgLeuLeuArgLysL 222
265   AAGATC... GCAGAACCGGAGATATCATTTGATGATGATGATGATGATCA 219
   |||||||||||||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
222   sArgSerPheGlnGlyLysGlnGlnGlyIleLeuAspValArgValGlnIle 239

```

```

218 ACACCCCTGAATAGCATGATGAATCAAGCTTTGCCAGTTGATTACGT 169
    |||||.....||| ||| |||
239 snthleuspiilewetglugly.....AlaserproserAlaproglu 253
    |||||.....||| ||| |||
168 CACCGGCACTTCTGCACATATGAGCAACCTGTATACTGCTGCAGT 119
    |||||.....||| ||| |||
254 AlaAlaAlaValSerIu.....GluserValaIargArgAlaTh 268
    |||||.....||| ||| |||
118 ACAAAATGCAATCTGCATGCAACATTTGGCTTCATGTTATGGGTT 69
    |||||.....||| ||| |||
268 rcerthnglyIleSerSerThProIlegIleuProserPheLeuglym 285
    |||||.....||| ||| |||
68 TAAGCATTCATTTGATTGCTGCCGTCATTATTACATTTGTTTAATT 19
    |||||.....||| ||| |||
285 eIaArgThrlIeValAlaThAlaLeuSerAlaThrlIePheYrVala 301
    |||||.....||| ||| |||
18 CMTCCA 13
    |||||.....||| ||| |||
302 ArgPro 303
    |||||.....||| ||| |||

seq_name: pIir2:SI7855
seq_documentation_block:
peptidylglycine monooxygenase (EC 1.14.17.3) - African clawed frog
N:Alternate names: peptidylhydroxyglycine N-C lyase
C:Species: Xenopus laevis (African clawed frog)
C>Date: 22-Nov-1993 #sequence_revision 03-Aug-1995 #text_change 18-Jun-1999
C:Accession: SI7855
R:Iwasaki, Y.; Kawahara, T.; Shimo, H.; Suzuki, K.; Ghisalba, O.; Kangawa, K.; Matsuo,
Eur. J. Biochem. 201, 551-559, 1991
A:Title: Purification and cDNA cloning of Xenopus laevis skin peptidylhydroxyglycine N-C
A:Reference number: SI7855; MUID:92037609
A:Accession: SI7855
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-935 <IWA>
A:Cross-references: EMBL:X62771; NID:664530; PIDN:CA44615.1; PID:664531
C:Superfamily: peptidylglycine monooxygenase II; peptidylglycine monooxygenase I homolog
C:Keywords: oxidoreductase
F:131-342/domain: peptidylglycine monooxygenase I homology <FCM>

alignment_scores:
      Quality: 100.00      Length: 391
      Ratio: 0.541      Gaps: 18
      Percent Similarity: 47.315      Percent Identity: 19.437

alignment_block:
US-09-323-427-5/rev x SI7855 ..
Align seg 1/1 to: SI7855 from: 1 to: 935

1065 TGTGACCACTTCAATTAACATCAATTTTAATACAGTAATGCAATGCA 1016
    ||| |||||.....||| ||| |||
526 CysGlnProThraspValaIaValasproIleThrGlyAsnPhphevea 542
    |||||.....||| ||| |||
1015 AGCAGATGTTATGTGAAGGCTTTATGATCAAGAAGGTGGCCGAATG 966
    |||||.....||| ||| |||
542 lAlaIaspGlyYr..... 546
    |||||.....||| ||| |||
965 ANGAAGTGAGCGTCAAGTGGCGAATTTCACTTCATTTGATTGATTCG 916
    |||||.....||| ||| |||
547 .....Cys 547
    |||||.....||| ||| |||
915 AATGTTGCCCGTCACAGATCTCTGAATCAGCTGGATTTTGTGA..... 871
    |||||.....||| ||| |||
548 Asn...SerArgIlelewetIinPheSerProAsnGlyMetPheIlewetG 563
    |||||.....||| ||| |||
871 ..... 871
    |||||.....||| ||| |||
563 nTrpGlyGluGluThnSerSerAsnValProArgProGlyGlnPheArgI 580
    |||||.....||| ||| |||
870 .....ACAACAACCTGTTGTCATTTGTTTCATCCATTAATTGTTACC 829
    |||||.....||| ||| |||

```

```

580 leProHISserleuThnMetValProAspGlnGlyInLeuCysValaIa 596
    |||||.....||| ||| |||
828 AAGTTGATGTCGATATGAGTACATGCTTTTACATGGAAGCTGATA 779
    |||||.....||| ||| |||
597 AspArgGluAsnGly...ArgIleGlnCysPheHISalaglThrGlyAs 612
    |||||.....||| ||| |||
778 AACAGTTAGTGCACAGATGAGTATCTGTAATC...ACAAGTGCSTTTC 732
    |||||.....||| ||| |||
612 nPheValLys...GlnIleLysHISGlnGluPheGlyArgGluValPhe 628
    |||||.....||| ||| |||
731 AAACCTAAATTTGTCGCCGATGCCAGTATGCCGTATGAATTTGGATGT 682
    |||||.....||| ||| |||
628 lValSerYrAlaPro.....Gly 634
    |||||.....||| ||| |||
681 GGACCAACCGGTCACACAGTTCATTTGCTATGATGTCAGGCACTTA 632
    |||||.....||| ||| |||
635 Gly.....ValleuYrAlaValaISnGlyLysProTyTr 646
    |||||.....||| ||| |||
631 TCATTAATGACATGGGATTCGAACCGTTGATFACTTTCTGGCGGTG 582
    |||||.....||| ||| |||
646 r.....GlyTyrSerAlaPro 652
    |||||.....||| ||| |||
581 TGCATTCCTGCTTGTTCGATGATGTAAGCGTATACCTGGAATTTCTA 532
    |||||.....||| ||| |||
652 alGlnIlePheMetLeuAsnPhSerAsnGlyAsp..... 663
    |||||.....||| ||| |||
531 AATGTCATGATGATGCTCTGTGATAATATTG.....CTAAATATT 488
    |||||.....||| ||| |||
664 .....IleLeuAspThrPheIleProAlaArgLysAsnPh 675
    |||||.....||| ||| |||
487 GGAATTTCCACAGATTTAATGCTGCGCCAGAGCTCAGTATACAAAT 438
    |||||.....||| ||| |||
675 eAspMetProHISAspIleAlaIaIaAspAspGlyThrValYrValaG 692
    |||||.....||| ||| |||
437 ATGCGGATGATCAGACGCTTTCTATCAATGCCAGATCAGT..... 397
    |||||.....||| ||| |||
692 lYAspAlaHISAlaAsnAlaValTrpLysPheSerProSerLysAlaGlu 708
    |||||.....||| ||| |||
396 ...ATTACCATTAAGAACCAAT.....ACGGAATGTCGTCAGCC 359
    |||||.....||| ||| |||
709 HISArgSerValLysLysAlaGlyIleGluValGlnGluIleThrGluTh 725
    |||||.....||| ||| |||
358 ACNATGTCAGAACACACAGATTCGAGCTGTTAAACAGGTCGCGCG 309
    |||||.....||| ||| |||
725 rGluIlePheGluThnHISIleArgSerArProLysThAsnGluSerV 742
    |||||.....||| ||| |||
308 CAGCAAAACCTGTCGAGCTGCGCAACTTCGTTACTCAAGAAAGATCT 259
    |||||.....||| ||| |||
742 alGluLysGlnThrGlnGluLysGlnGlnLys.....GlnLysAsnSer 756
    |||||.....||| ||| |||
258 GCA.....GACCGGAGATATTCATTT...GATGTCAGAAC 227
    |||||.....||| ||| |||
757 AlaglYvalSerThrGlnGluLysGlnAsnValaIaGlnGluIleAsnAl 773
    |||||.....||| ||| |||
226 TGATATCAACACCCCTGAA.....ATTAGCGTGAATATCAAGCTT 186
    |||||.....||| ||| |||
773 acIYvalProThnGlnGluLysGlnAsnValaIaGlnGluSerAlaG 790
    |||||.....||| ||| |||
185 TGCACTGATTACGTCACCGTCGCACTTCGCAACATATGACCAACT 136
    |||||.....||| ||| |||
790 lYvalSerThrGlnGluLysGlnSerValaIaGlnGluSerAlaGly 806
    |||||.....||| ||| |||
135 GTAATATCTTCGTCAGTACAAAATGCAATCTGCATTCACATTTGCGTT 86
    |||||.....||| ||| |||
807 ValSerThrGlnGluLysGlnSerValaIaGlnGluSerAlaIaY 823
    |||||.....||| ||| |||
85 CTCATGTTATGGGTTTAAGCATTCATTTGATTT...GCTGCCGTATTA 39
    |||||.....||| ||| |||
823 lSerPheValIleIleIleThrLeuLeuIleIleProIleAlaValLeu 840
    |||||.....||| ||| |||
38 TTACCATTTGTTTAATTTGCT 16
    |||||.....||| ||| |||

```


Align seg 1/1 to: C70126 from: 1 to: 610

```

1146 ATTGCTTCTGACTACACTATTCGATTG..... 1117
|||||.....
103 IleAlaIleCysSerAsnIleSerIleThrSerSerThrIleSerAsnGlu 119
|||||.....
1116 TCTATTTCGATTCCGGTTCGACATGTCGCA.....GGTG 1080
|||||.....
119 uSerIleGlnIleGluValGluAsnGlyIleGluIleCysPheIleLysG 136
|||||.....
1079 AGCCAGAAATGAA.....TGTGACCACTTCATTAACATCAAT 1039
|||||.....
136 InProAlaIleAsnGlyThrIleValAlaIlePheIleThrIlePheHisAsn 152
|||||.....
1038 TTTAATACAGCTTAATGACATTCGACAGACATGTTTATGTCGAAGGCTTTA 989
|||||.....
153 PheProAlaIleArgLysArgPheLeuLysGlnIleProIle.GluThrLysM 169
|||||.....
988 TGATCAAGAGAGGTGCCGTAATGATGAGAGGTGACGTCAGATTGCCGGA 939
|||||.....
169 eCysLeuLys.....ValLeuGluGluLysIleIleThrHisProGlu 183
|||||.....
938 TTTCACTTCATTTGATTGATTCGCAATGTTGCCGTCACAGATCTGCAAT 889
|||||.....
184 IleAsnPheGlu..... 187
|||||.....
888 CCACGTGGTATTTTGTACACAACTGTTGTCATTTGTTGTTCAATCCATT 839
|||||.....
188 .....IleAsnLeuAsnGlnLysLeuArgLysIleTyrPheLysGlu 202
|||||.....
838 ATTGTTCACCAAGTTG..ATCGTGCAAT.....ATCGAGTACCAAT 801
|||||.....
202 eLeuIleAspArgValGlnAsnValTyrGlnLysValIleGluAsnAsn 218
|||||.....
800 GCTTTTACATGGAAGCTGATAAACAAGTTAGTCACAGATGAGGTATCT 751
|||||.....
219 LysPhe.....ArgValIle 223
|||||.....
750 GAATTCACAA.....CAGCTTTCAACACTCA 725
|||||.....
223 uLysLysGlnHisAspAsnIleLysIleGluIlePheLeuAlaProAspA 240
|||||.....
724 AATTGTCCCGATGCGAGTATGCCGTTATGAATTTTGATGGTGGACCAA 675
|||||.....
240 snPheSerLysSerLysArgHisIleLysThrPheValAsnArgArg 256
|||||.....
674 CCGGTCACCACTCAATTTGCTATCATTTGTCAGCCAGTTATCATATAA 625
|||||.....
257 ProIleAspGlnLysAspLeu.....GluAlaIleThrAs 269
|||||.....
624 TGGACATGGG...ATTCTGAACCGTTGATCTTCTGCGCGGTCTGCA 578
|||||.....
269 nCylHisSerArgIleLeuSerPro..... 277
|||||.....
577 TTCGCTTTTGTGATGATGTACGGTGTACTGTAATTCCTTAATG 528
|||||.....
278 .....GlyAsnPheProIle 282
|||||.....
527 CTGATGATGTCCTCTGTATAATAATTTGCTAATAATTTGGAATAATCA 478
|||||.....
283 CysTyrLeuPheLeuGlnIleAsnProGluTyrIleAspPheAsnValHi 299
|||||.....
477 ACAGATTTAATGGCTGGCCAGACAGCTCACGTAATACAAATATGCGGATG 428
|||||.....
299 sPro.....GlnLysLysGluValAlaArgPheTyrAsn..... 309
|||||.....
427 ATCACAGCTTTCTATCAATGCCAGATCAGTATTAACA..... 390
|||||.....
310 .....LeuProPheLeuPheLys 315
|||||.....
389 ...TTAAAGAACCAATAGCGAATGTGTTGACACCAATGTTTCAGAACCA 343
|||||.....

```

```

316 LeuIleSerAspAsnIleAsnAsnPheAspLysAsnIleAsnAsnTyr 332
342 CAAGCATTCGAGCTGTTTAAACAGGTGGTGGCCGACGAAACCTGCTGC 293
|||||.....
332 rGlnAspIleIleIleLysArgGlnLeuThrGlnAspAsnIleIleG 349
|||||.....
292 AGCTGGCAGCACTTCGTTTACTCAAGAAAGATCTGCAGAACGAGATA 243
|||||.....
349 IuMetThrAsn.....Gln..ProGluAsnP 357
|||||.....
242 TCATTGATGATACGAACCT..GATATC.....AACACCTTGAATT 205
|||||.....
357 heAsnLysIleAsnThrTyrAspIleProGlnAsnAsnLeuGluThr 373
|||||.....
204 AGCGATGATATCA.....GCTTCCAGTGTGATT 173
|||||.....
374 GluAspValAsnGluProAsnLysAsnThrThrGlnSerAsnIleAsp 390
|||||.....
172 A...CGTCAACGTCGACTTCTGCAACATTAATGACAACTGTAATACTG 126
|||||.....
390 uArgArgTyrAsnSerIleIleGlnAsn.....ArgProThrLeuArg 405
|||||.....
125 CTGCACTACAAAT 112
|||||.....
405 LuAsnIleGluAsn 409
|||||.....

seq_name: p1r2:T11616

seq_documentation_block:
carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5) - fission yeast (Sc
C/Species: Schizosaccharomyces pombe
C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
R:Barrell, B.G.; Raeburn, M.A.; Walsh, S.V.; Wood, V.
submitted to the EMBL Data Library, October 1995
A:Reference number: Z17300
A:Accession: T11616
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-224 <BAR>
A:Cross-references: EMBL:Z54328; NID:91009451; PID:91009456
A:Experimental source: strain 972h(-)
C/Genetics:
A:Map position: IL
A>Note: SPAC2267.06c
C:Keywords: ligase

alignment_scores:
Quality: 92.00 Length: 365
Ratio: 0.544 Gaps: 17
Percent Similarity: 46.301 Percent Identity: 20.274

alignment_block:
US-09-323-427-5/rev x T11616 ..
Align seg 1/1 to: T11616 from: 1 to: 2244

1038 TTTAATACAGTAATGCAATTCGACAGACATGTTTATGGAAGGCTTTA 989
|||||.....
999 TyrThrThrTyrAsnAlaValGlnHisAspIleHisPhe..... 1011
|||||.....
988 TGATCAAGAAGGTGCCGTAATGATCAAGCT.....*.....G 957
|||||.....
1012 .....AsnAspLysGluValMetValLeuGlySerG 1022
|||||.....
956 GACGTCAAAGTTCGCGAATTTTCATTCATTTGATTCATTCATTCGCG 907
|||||.....
1022 TyValTyrArgIleGlySerSerValGluPheAspTyrCysAlaValArg 1038
|||||.....
906 CGTACAGATCTCTGAATCCAGTGGTATTTTGTAAACAACAGTGTGT 857
|||||.....
1039 AlaValArgThrLeuArgAspArgGlyVal.....LysThrIleLeu 1052
|||||.....

```



```

354 rpfhrAlaasnCysHslyscysfhrCysfhrAspAlaIuThrValasp 370
597 .....ACTTCTGCGCGGTTGTCATTCCTGCTTGTGGANGG 557
371 CysLysLeuLysGluCysProserProPthrCysLysProGluu... 386
556 TAACGGTGATCTGTGAATTCATAATGCTGATGATGCTGCTTGATA 507
387 .....ArgLeuValLysPheLysAspAsnAspThrCysGluIleA 401
506 AATAT.....TTGCTAATAATATTGGAAATCCACA 475
401 IatYcysGluProArgThrCysLeuPheAsnAsnAspTyr..... 415
474 GATTAAATGCTGGCCAGAACCTCAGTATCAATAATATGCGAGTCATC 425
415 ..... 415
424 ACAGCTTTTCTATCAATGCCAGATAGATTATTCATTAAGAACCAATA 375
416 .....GluValGlyAlaSerPheAlaAspProLysA 426
374 GCGAATGTGTGACCAATCTTCAGAACCAAGAGATTGCGAGCTGT 325
426 snProCysIleSerTyrSerCysHsAsn...ThrGlyPheValAlaVal 441
324 AA.....AACAGTGTGTCCGC..... 308
442 ValGlnAspCysProLysGlnThrTrpCysAlaGluGlnAspArgValTy 458
307 .....AGCAAAACCTGCTGCAGCTGGCAACTTCGTTTACTCAAGAAAGA 262
458 rAspSerThrLysCysCys.....TyrThrCysLysProT 470
261 TTCGACGAACCGAGAAATATCATTTGATGACGAC 227
470 YrcysArgSerSerSerValAsnValThrValAsn 481

```

seq_name: plr2:S72278

seq_documentation_block:
 ATP-dependent Clp proteinase (EC 3.4.21.-) homolog - Plasmodium falciparum plastid
 C:Species: plastid Plasmodium falciparum
 C:Date: 14-Apr-1998 #sequence_revision 24-Apr-1998 #text_change 07-May-1999
 C:Accession: S72278: S78483
 R:Wilson, R.J.M.; Denny, P.W.; Preiser, P.R.; Rangachari, K.; Roberts, K.; Roy, A.; Whyt
 J.; Mol. Biol. 261, 155-172, 1996
 A:Title: Complete gene map of the plastid-like DNA of the malaria parasite Plasmodium fa
 A:Reference number: S72277; MUID:96346169
 A:Accession: S72278
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-765 <WIL>
 A:Cross-references: EMBL:X95276
 R:Wilson, R.J.M.
 submitted to the EMBL Data Library, January 1996
 A:Reference number: S78483
 A:Accession: S78483
 A:Molecule type: DNA
 A:Residues: 1-95, 'N', 96-765 <WIL>
 A:Cross-references: EMBL:X95276; NID:g1171591; PID:e220210; PID:g1171612
 C:Genetics:
 A:Gene: clpC
 A:Genome: plastid
 A:Note: this apparently degenerate plastid is referred to as the apicoplast
 C:Keywords: hydrolase; plastid; serine proteinase

alignment_scores: Quality: 90.00 Length: 225
 Ratio: 0.720 Gaps: 20
 Percent Similarity: 55.356 Percent Identity: 27.111

alignment_block:
 US-09-323-427-5/rev x S72278
 Align seg 1/1 to: S72278 from: 1 to: 765

```

848 TTCATTCATTAATTGTTTACCAAGTTGATC.....GTGCATATC.. 810
282 TyrLeuPheLeuLeuLeuAsnLysLeuTyrGlyTyrAsnIleHsIleI 298
809 .....GAGTACATGCT...TTTACATGGAAGCTGA 782
298 eIleValThrAsnLysGluIuTrpAsnThrTyrPheLysTyrAsnIleI 315
781 TAAACAG.....TTAGTCACAGATGAGTATCG 750
315 IeLysAspSerTyrPheTyrLysIleArgIleLysAspLeuIleLeu 331
749 AAATCAACAACCTGTTTCAACTCAATTTGTCGATGCCAGTATGCCGT 700
332 GlnThrPheLeuIleIleLysAsnAsnIle.....TyrLysTy 344
699 TATCAAAATTTTGGATGCTGGACCAACGGTCAACAG.....TTCA 659
344 rIleAsnTyrTyrLysIleAsn.....IleAsnAsnTyrIleIleTyrG 359
658 ATTGCTATCATGTCGACGCCAGTTTATCATTAATGACATGCCATTCTG 609
359 IuLeuIleAsnLeuSerLysTyrIle..... 368
608 AAACGTTGATACTTCTGCGCGGTTGCATTCCTGCTTTCGATGAT 559
369 LysProLeuIleLeu.....ProThrThrProLeu.. 378
558 GGTAAACGCTGATCTGTGGAATTTCAATGCTGATGATGCTGCTTGA 509
379 .....IleLeuLeu.GluAsnSer.....CysSerAs 387
508 TAAATATTGCTAAATTAAT...TTGGAATATCCACACAGATTAAAGCTG 462
387 nLysTyrLeuLeuAsnAsnLysIleSerTyr.SerAsnAsnAsnTyrLeu 403
461 GCCAAGAACCTCAGTATACAAATATGCGATGATCACAGCTTTCTAT 412
404 .....PheThrTyrAsnAsn..AsnIleIleTyrAsnAsnLysAs 416
411 CATGCGCATCAGTATTAACCAATTAAGAACCAATAGCGAATGTGTCG 362
416 nAsnAsnLeuThrIleGluAspIleLysAsnSerIleSerAsnTyrLeu. 432
361 ACCACAAATGTTCAAGAC.....CACAAAGATTGCGAGCTGTAAAA 321
433 .....AsnIleSerLysThrIleLeuPheLysAspAsnLysLeuThrLys 447
320 CAGGTGTGCGCCAGCAAAACCTGCTGACAGTGCACCAATTCGTTTACTC 271
448 .....LeuAsnLeuThrLysLeuGluAsnTyrLeuThrAs 459
270 AAGAAAGATCTGCAGAC 252
459 nHsIleTyrGlyGlnAsn 465

```

seq_name: plr2:S48944

seq_documentation_block:
 hypothetical protein YHR102w - yeast (Saccharomyces cerevisiae)
 C:Species: Saccharomyces cerevisiae
 C:Date: 13-Jan-1995 #sequence_revision 10-Feb-1995 #text_change 24-Sep-1999
 C:Accession: S48944
 R:Latelle, P.
 submitted to the EMBL Data Library, May 1994
 A:Description: The sequence of S. cerevisiae cosmid 8263.
 A:Reference number: S46691

A:Accession: S48944
 A:Molecule type: DNA
 A:Residues: 1-1080 <LAT>
 A:Cross-references: EMBL:U00059; NID:g529116; PIDN:AA68860.1; PID:g529127; MIPS:YHR102w
 C:Genetics:
 A:Gene: SGD:NRK1
 A:Cross-references: SGD:S0001144; MIPS:YHR102w
 A:Map position: 8R
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
 C:Keywords: ATP: serine/threonine-specific protein kinase
 F:21-276/Domain: protein kinase homology <KIN>
 F:29-37/Region: protein kinase ATP-binding motif

alignment_scores:
 Quality: 90.00 Length: 209
 Ratio: 0.833 Gaps: 6
 Percent Similarity: 51.675 Percent Identity: 21.531

alignment_block:
 US-09-323-427-5/rev x S48944 ..

Align seg 1/1 to: S48944 from: 1 to: 1080

```

732 CAAACTCAAAATTGTCCTCCGATGCCGATATGCCGTTATGAATTTGGATGG 663
      :::::::::::  :::::::::::  |||
200 LysValAspIleThrSerLeuGlyIleThrThrGluIleAlaThrGlu 216
      :::::::::::  :::::::::::  :::::::::::
682 TGGACCA.....ACCGTCAACCAAGTTTCATTTGCTA 651
      :::::::::::  :::::::::::  :::::::::::
216 yAsnProProTyrCysAspValGluAlaLeuArgAlaMetGluLeuIle 233
      :::::::::::  :::::::::::  :::::::::::
650 TCATTTGGTCAGCCGATTTCATTAATGCATGCATGCTTGTGAACCGTT 601
      || :::::::::::  |||
233 IeYSerLeYsProProArgLeuGluAspArgSerLeYSerThrSerLeu 249
      :::::::::::  :::::::::::  :::::::::::
600 GATACCTTTGCCCGGTTGCCATTCCTGCTTTGCGATGATGTAACG 551
      :::::::::::  :::::::::::  :::::::::::
250 LysGluPheIleAlaLeu.....CysLeuAspGluAspProlys.. 262
      :::::::::::  :::::::::::  :::::::::::
550 TGATCTGTGGAAATCTTAATGCTGATGATGCTGCTGATTAATAT 501
      ||| ||| :::::::::::  |||
263 .....GluArgLeuSerAlaAspSerLeuLeuLysSerLeYsPheI 216
      :::::::::::  :::::::::::  :::::::::::
500 TGCTAAATATTTGGAATATCCACAGATTTAATGGCTGGCCCAAGAGCT 451
      :::::::::::  :::::::::::  :::::::::::
276 IeArgAlaHisLysAlaThrProThrSerLeuLeuLysGluLeuIleSer 292
      :::::::::::  :::::::::::  :::::::::::
450 CACGATATACAAATATGCCGATGCATCACAGCTTTCTATCATATGCCAGAT 401
      :::::::::::  :::::::::::  :::::::::::
293 ArgTyrLeuLeuPheArgAspLysAsnLysAsnLysTyrLysIleGlu 309
      :::::::::::  :::::::::::  :::::::::::
400 CAGTATACCATTAAGAACCAATAGCGAATGTGTCGACACCAATGTT 351
      ||| ||| :::::::::::  |||
309 ySerIleProGluAsnGluProSerLysProSerGluAlaProLysProS 326
      :::::::::::  :::::::::::  :::::::::::
350 CAGAACCAACAGAGATTGCGAGCTGTTAAACAGTGTCGCCGACAGAAA 301
      :::::::::::  :::::::::::  :::::::::::
326 eArgL.....AsnGlyGlyLysAspGluAlaGln 335
      :::::::::::  :::::::::::  :::::::::::
300 CCTGCTGCAGCTGCCAAGCTGCTTACTACAGAAAAGATCTGCAGAAC 251
      :::::::::::  :::::::::::  :::::::::::
336 LysSerIleLeuAsnAspAsnGluIleLysArgValAsn..... 349
      :::::::::::  :::::::::::  :::::::::::
250 GAGAGATATCATGATGATGACGATGATATCACACCTTGAATATGCG 201
      ||| ||| :::::::::::  |||
350 .GluGlyAspValGluMetLysTyrAspPheAspSerLeuSerSerSera 366
      :::::::::::  :::::::::::  :::::::::::
200 AT.....GATATCAACGCTTTGCCAGTTGATTTA 172
      :::::::::::  :::::::::::  :::::::::::
366 spYrIleIleGluAsnAsnIleAsnLeuAspAlaLeuAlaGluAspAsn 382
      :::::::::::  :::::::::::  :::::::::::
171 CGTCAACGCTGCACCTTCTGCAACATAAT 145
      :::::::::::  :::::::::::  :::::::::::

```

seq_name: plr2:C71618
 383 AsnGluTrpAlaThrAlaGlnHisAsp 391

seq_documentation_block:
 Hypoetical protein PFB0315w - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 07-May-1999
 C:Accession: C71618
 R:Gardner, M.J.; Tellez, H.; Carrico, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.; Perera, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H. Science 282, 1126-1132, 1998
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference number: A71600; MUID:99021743
 A:Accession: C71618
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1712 <GAR>
 A:Cross-references: GB:AE001386; GB:AE001362; NID:g3845148; PID:g3845151; TIGR:PFB031
 A:Experimental source: clone 3D7
 C:Genetics:
 A:Gene: PFB0315w

alignment_scores:
 Quality: 90.00 Length: 341
 Ratio: 0.584 Gaps: 12
 Percent Similarity: 45.161 Percent Identity: 19.941

alignment_block:
 US-09-323-427-5/rev x C71618 ..

Align seg 1/1 to: C71618 from: 1 to: 1712

```

989 ARGATCAAGAAGGTCCCGTAATGATGAAGCTGACGATCAAGTTGCCGA 940
      ||| ||| ||| :::::::::::  |||
1102 MetLysLysLysLysThrGlnGluTyrValAspIleGlu.....Th 1116
      :::::::::::  :::::::::::  :::::::::::
939 ATTTCACCTTCATTTGATTCATGCATGTTGCCGCTACAGATCTCGAA 890
      :::::::::::  :::::::::::  :::::::::::
1116 rValTyrGluTyrValIleGluLysTyrLysArgValHisValLeuTyrL 1133
      :::::::::::  :::::::::::  :::::::::::
889 TC...CACGTGATTTTGTGTAACAACAACACTGTTCATTTGCTTCATC 843
      :::::::::::  :::::::::::  :::::::::::
1133 eudLArgLeuLeuGlnIleValGluLysLeuPheLysLysTyrIleLeu 1149
      :::::::::::  :::::::::::  :::::::::::
842 CATTATTTGTTACCAAGTTGATCGTCGATATCGATACATGCTTTTAC 793
      :::::::::::  :::::::::::  :::::::::::
1150 LysTyrSerPheHisLysLeuArgIlePheTyrGluTyrLysIleGlu 1166
      :::::::::::  :::::::::::  :::::::::::
792 ATGGACGCTGATTAACAGCTAGTCACAGATTGAGGTATCGAATTCAC 743
      ||| ||| ||| :::::::::::  |||
1166 tGluLysLeuLysLysAsnTyrIleHisCysIleTyrAspIleSerAspL 1183
      :::::::::::  :::::::::::  :::::::::::
742 AACCTGTTTCAACTCAATTAATGTCGCCGATGCCAGTATGCCGTTATGAAA 693
      :::::::::::  :::::::::::  :::::::::::
1183 ySLeuGluPheLeuIleLysLysLysMetGlnHisTyr..... 1195
      :::::::::::  :::::::::::  :::::::::::
692 TTTTGATGTGTGACCAACCGGTCAACCAAGTTCAATTTGCTATCATTTGGT 643
      :::::::::::  :::::::::::  :::::::::::
1196 .....PheAsnHisIleIleIleAsnSerTyrGln 1205
      :::::::::::  :::::::::::  :::::::::::
642 CACGAGTTTATC.....ATAATGAGCATGCGATTGAAACCGT 602
      ||| ||| ||| :::::::::::  |||
1205 uSerSerPheIleAsnTyrGlnIleLysThrAsnAspMetLeuTyrAsnL 1222
      :::::::::::  :::::::::::  :::::::::::
601 TGATACCTTCTGCGCGGTGTCATTCCT...GCTTGTGATGATGATA 555
      :::::::::::  :::::::::::  :::::::::::
1222 euleuLeuLysGluLysSerAlaTyrGlnAsnHisLeuGluLysAsnTyr 1238
      :::::::::::  :::::::::::  :::::::::::
554 ACCGTATATCTGTGGAATTTCAATGCTGATG.....GATGTGCTCTT 511
      :::::::::::  :::::::::::  :::::::::::

```

```

1239 IleuIleuIleuTyrIysValIleuLeuSerMetTyrIysLysMetAl 1255
510 GATAAATATTGCTAATAATTGGAATATCAACAGATTTAATGCGTGG 461
||| ::::::::::||| :::::
1255 alleTyrPheArgSerPheValItyrAsnIleIysValSerLysLysL 1272
460 CCAAGAGCTCACGTATACAAATATGCGATGCATCAC..... 423
::: ::::::::::||| ||:::
1272 ysaSnIaPheAlaTyrThrLeuThrArgValAsnSerIleLeuVal 1288
423 ..... 423
1289 TyrGlutArgArgIleIysSerPheIlePheSerLysLeuLysPheAsnTy 1305
422 .....AGCTTTTCATCAATGCCAGAT 401
1305 rasPaSnValSerTyrPheCysPheThrMetTyrIysIleTyrLeuArgA 1322
400 CAGTATTA.....CCATTAAAGAACCAATAGCG 372
:::|::| ::::|::|
1322 rglLeuPheGlyTyrLeuArgIleArgPaSnArgIleAsnIleLys 1338
371 AATGTGTGCACCAATGTT..... 351
|||::|::|::|::|::|::|
1339 AsnValIleGluIysAsnValTyrArgLeuValLysLeuIleSerLysI 1355
350 ....CAGAACCACACAGATTCGAGCTGTTAAACAGGTGTGCCGACG 305
:::|::|::|::|::|::|::|
1355 eSerAspAsnHisLysTyrAsnIaPheLeuLys..... 1366
304 AAAACCTGCTGCAGCTGCCGCACTGTTTACTCAAGAAAAAGATCTGCAG 255
|||::|::|::|::|::|::|::|
1367 .....LeuGlnLysTyrValTyrGluGlnAsnGluLysLys 1378
254 AACCGAGAGATATCATTTGATGTACGAACTGATCAACACCTTGAATT 205
|||::|::|
1379 AsnLysMetLe..... 1382
204 AGCGATGATATCAACGCTTTC...CAGTTGATTACGTCACCGTCACAT 158
||| ::::|::|::|::|::|::|::|
1383 .....CysAspAsnLeuIleTyrAlaAsnGlnL 1393
157 TCTGCAACATATGACAACTCG 135
::|::|::|::|::|::|::|
1393 eucYsaSnAsnLeuAspLysIle 1400

seq_name: pir2:A41519

seq_documentation_block:
posterior-group protein tudor - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 24-Sep-1999
C:Accession: A41519; S19019
R:Golumbeski, G.S., Bardsley, A.: Tax, F.: Boswell, R.E.
Genes Dev. 5, 2060-2070, 1991
A>Title: tudor, a posterior-group gene of Drosophila melanogaster, encodes a novel prote
A:Reference number: A41519; MUID:92038995
A:Accession: A41519
A:Molecule type: mRNA
A:Residues: 1-2515 <GOL>
A:Cross-references: GB:X62420; NID:98753; PIDN:CAAA4286.1; PID:98754
A:Gene: tud
C:Genetics:
A:Gene: tud
A:Cross-references: FlyBase:FBgn0003891
C:Superfamily: posterior-group protein tudor

alignment_scores:
Quality: 90.00 Length: 367
Ratio: 0.552 Gaps: 16
Percent Similarity: 44.414 Percent Identity: 19.074

alignment_block:

```

```

US-09-323-427-5/rev x A41519 ..
Align seg 1/1 to: A41519 from: 1 to: 2515
1116 TCTTATTCGATTCGCGGTGACAAATGCTGCAAGGTGAGCCGAAATTGA 1067
||| ::::|::|::|::|::|::|::|
1904 SerLeuGlnLeuPro...AspAlaTyrIleSerTrpSerProGluIaGl 1919
1066 ATGTGGACCACTCAATACA.....ATCAATTTTAATACAC 1029
|||:::|::|::|::|::|::|::|
1919 uAlaLysPheAlaGluLeuThrGlyGluValPheThrThrg 1936
1028 GTAATGCATTCGAAGA.....CATGTTTATGCAAGCTCTTATGAT 985
::|::|::|::|::|::|::|::|
1936 lIleuLeuLysProGlyGlnAspHisValThrIleAspLeuLeuAsp 1952
984 CAAAGAGGTGGCCGTATATGATGAGGTGACGTCAAGTGGCCGAATTTC 935
|||:::|::|::|::|::|::|::|
1953 .....GlyGluAsnIleIleAspArgLe 1960
934 ACTTCATTTGATTCATGCAATGTCGCGTACAGATCTGTAATCCAC 885
|||::|::|::|::|::|::|::|
1960 uLeuProLeu....CysGlnArgLysGluProLysGluAlaSerLysG 1975
884 GTGGATATTTTGTACACACACTGTGTGATTCATTCATTCATTAATT 835
|||::|::|::|::|::|::|::|
1975 lSerIleuAlaValThrThrLysAlaIleIle..... 1985
834 GTTACCAAAAGTTGATGTCGATTCGATTCGATTCGATTCGATTCGATTC 785
|||::|::|::|::|::|::|::|
1986 ...ThrHisValGlnAsnThrSerArgIleTyrLeuGlnPheSerGlu 2001
784 TGAT..... 781
2001 sasPserLeuMetAspIleIleCysGluLysLeuAsnGlySerLysLeuG 2018
|||::|::|
780 ....AAACAGTTAGTCGACAGATTGAG..... 757
|||::|::|::|::|::|::|
2018 lnrProLysThrGluLysAlaIleValAspAspMetCysValAlaGlnPhe 2034
756 .....GTAATCGAATCAACAATCTGC 737
2035 AlaAspAspLeuGluPheTyrArgSerArgIleLeuGlnValLeuGlnAs 2051
736 TTTTCAACATCAATATGTCGCCATGTCAGATGCCCTTATGAATTTGG 687
|||::|::|::|::|::|::|
2051 pasPrlnTyrLysValIle.....LeuIleA 2060
686 ATGGTGGACCAACC.....GGTCAACCACTTCA 658
|||::|::|
2060 sPTyrGlnAsnThrThrValAlaLysPheLysLeuTyrGluLeuProGlnGlu 2076
657 TTTGCTATCATTTGTCAGCCAGTTTATCAATAATGACATGCGATTTCTGA 608
|||::|::|::|::|::|::|::|
2077 PheThrLeuIle...LysProValAlaGlu.....IleCysSerMetG 2090
607 A.....ACCGTTGATACCTTCT 591
2090 uProSerAlaIlePheGluLysAsnLysAlaLeuThrIleThrPheA 2107
|||::|::|::|::|::|
2107 spAlaLeuLeuAspSerCys..... 2113
540 GAAATTTCTAATGCTGATGATGCTCTTGAATAATATTGCTTAATTA 491
|||::|::|::|::|::|::|
2114 .....LysGlyValValAlaValAlaGluPheValAsnLysSe 2125
490 TTTGAAATATCCACAGATTTAATGCTGGCCAAAGAGCTCCAGTATACA 441
|||::|::|::|::|::|::|
2125 rAlaSerProProValValArgLeuThrThrLysAspLysArgSerLeu 2142
440 AATATCGGATGCATCAGAGCTTTCTATCAATGCCAGATCAGTATTACC 391

```

```

11 ..... ||| :: |||.....
2142 ysllerytgcunhslsleuglnlyslsleuvalglmlaiglnleuylsleuile 2158
390 ATTTAA... GAACCAATAGCGAATGCTTGCAGCAACATGTTCCAGAAC 344
||| :: |||.....
2159 glnlysrtrgsnclunbsnserglucyslleilserlytrclysnsrpr 2175
343 ACAAGATTCCGAGCTGTTAAACAGGTGGCCGACAAACCTGCTG 294
||||| :: |||.....
2175 olusserphetryValglmlslyshsasnserrAlaAspleuAspleuI 2192
293 CACCTGGCAGCACTTCTTTCCTCAAGAAAGATCTGCAGAACCGAGAT 244
||||| :: |||.....
2192 leVallysthrleuglnserleuylslysglnlyslsleu.....Lyslys 2206
243 ATCATTGATTAGCACTGATATCACACCTTGAATTAGCGATGATTA 194
||||| :: |||.....
2207 leuileasprrothrtthrasnsrsglyvalcystlyserglnluAs 2223
193 T 193
2223 p 2223

seq_name: pir2:H69898

seq_documentation_block:
hypotheoretical protein yobI - Bacillus subtilis
C.Species: Bacillus subtilis
C.Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1998
C.Accession: H69898
R.Kunst, F.; Ogatawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertei
C.; Bron, S.; Brouillet, S.; Bursch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.; Chc
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
lech, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koelter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinols,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y. M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivoalta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon
A:Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiyuchi, J.; Sekowska, A.; Serot
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumschein, E.; Yoshikawa, H.; Danchin, A.
A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033
A:Accession: H69898
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1201 <KUN>
A:Cross-references: GB:299114; GB:AL009126; NID:92634230; PID:el185369; PID:g2634290
A:Experimental source: strain 168
C:Genetics:
A:Gene: yobI

alignment_scores:
Quality: 89.50 Length: 392
Ratio: 0.511 Gaps: 17
Percent Similarity: 44.643 Percent Identity: 18.367

alignment_block:
US-09-323-427-rev x H69898 ..
Align seg 1/1 to: H69898 from: 1 to: 1201

1100 TTGACAATGCTGTCGAAGTGCAGCAAAATGATGTGACCACTTCA 1051
||| :: |||.....
552 leuileuglnleuglnuarglnlysglnluileargAlaGlnserleuGI 568
1050 ATTACACATCAATTTTAATACAGCTAATGCATTGGAAGACACATGTTATGT 1001
| :: |||.....
568 n.....ValleuilerhrlsyrmetasnsrlyssAspvalphesera 582

```

```

1000 GAAAGCTTTTATGATCAAGACAGTTGCCCTAATGATGAAGTGCAGCTC 951
||||| :: |||.....
582 splysleuylrglu...LyslysleuileuVallyrleuileuAlghnlsGly 597
950 AAGTGGCCGAATTCACCTTCCATTGATT..... 921
::: |||.....
598 tyrlleasprlmetuylrasnhtslyrllrthrtphetryrprogluse 614
920 ..... CATGCA 914
614 rleuSerleuSerAsprlleLysPheValPheSerlleLysAsnhtsGluS 631
913 TGTG..... CCGCTACACGATCTC 894
631 erleuProtyrSerPheGlnleuAspAsnllleGlyLyslleMetserLys 647
893 TGAATCCAGCTGCTATTTTGTAAACAACACTGTGTCATTGCTTTCAT 844
::: :: |||.....
648 leuValcglylaGlnleuPheLysGlnlleGluValleuAsnPhenhtsleu 664
843 CCAATTATTTGTTACCAAGTTGATCGTCATGCAATGCAATGCTTTTA 794
664 uAsnTyrlleMetAsp.....HisserGluTytrArgAsnTytr 677
793 CATGAGACTGATTAACAAGTTAGTCAC..... 765
677 yrAspSerllelleGlnluargleuAlaAsnclYserLysGlnserValThr 653
765 ..... 765
694 PheilleasprlyPheLysGlnuargAlaAlaAsnlyAlaAlaPheilleG 710
765 ..... 765
710 nserlleSerSerLysTytrAspAspPheTytrPheilleGlnleuArgS 727
764 ..... AGATTGAGTATGGAATCACAAGCTTT 735
727 erAsnTytrThrGlnGlnLysLysGlnGlnLysleuSerAsprlleuThr 743
734 TTCAAA.....CTCAATTGTCGCCGATGCCAGTATGCCGT 700
::: :: |||.....
744 TyrlAlaAsprlleAlaAsprllelleArgMetAsnLys.....GluSerVa 758
699 TATGAATTTTGGATGTGAGCAACCGGTACACAGTTCAATTGCTCTAT 650
758 lmetSerPheThrleu.....SerLysTytrleuAsnleuS 771
649 CATTTGTCAGCCAGTTTATCATTAATGCACATGCGATTTCGAACCGTTG 600
||||| :: |||.....
771 erleuValSerAsp.....GlnluLyslleLysGlnleu 782
599 ATACTTTCGCGCGGTTCATCTCTGCTTGTGC...ATGATGTAAC 553
::: |||.....
783 leuileuLysleuclnValLysPheLysSerleuAsphtsleuSns 799
552 GCTGATACGTGGAATTCATAATGCTGATGATGATGCTGCTGATATAA 503
799 rGlnuThrlleryrAspPheValAlaGlnArgAsnleuTyrglnldeAsn 816
502 TT.....TGCTAAATTAATTTGGAATATCCACAGATTTAATGGCTGGC 459
||| :: |||.....
816 leLysThrleuSerVallleleuAsnAsp.....Ala 826
458 AAGACCTACAGTATACAAATATGCGGATGCATCAGACCTTTTCATCAA 409
::: |||.....
827 ProAsnllrthrtlyrAlaAlaVallyLysAsnserAsp..... 838
408 TGCCACATCAGATTTCATTAAGAACCAATAGCAATGCTGTTGAGAC 359
839 .....GlnGlnAlaVallleAsnTytrValAsnA 848

```


THIS PAGE BLANK (USPTO)

OM of: US-09-323-427-5 to: SwissProt_38:* out_format : pfs

Date: Apr 16, 2000 5:00 AM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+ncp.model -DEV=xip
-O=/cgn2.1/USPTO.spool/US09323427/runt.14042000_170514_1993/app_query.fasta.1
-DB=SwissProt_38 -QEXT=fasten -SUFFIX=backtrans.rsp
-GAPOP=12.000 -GAEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -GAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000
-XGAEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-YGAEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=bloms62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
-THR_SCORE=pct -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
-MINLEN=0 -MAXLEN=1000000 -USER=US09323427 -NCPU=6 -ICPU=3
-NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-323-427-5
Query length: 1161
Database: SwissProt_38:*
Database sequences: 82229
Database length: 29864866
Search time (sec): 138.700000

score_list:

Sequence	Strd Orig	Zscore	Escore	Len	Documentation
SwissProt_38:V33 CAEEL -	1420.50	2304.43	2.1e-121	424	003755 caenorhabditis elegans
SwissProt_38:Y0H3 CAEEL -	108.00	159.52	0.0538	495	009276 caenorhabditis elegans
SwissProt_38:YVY3 SCCHO -	97.50	139.76	0.5168	649	010211 schistosoma haematophyllum
SwissProt_38:AMU0 PIG -	97.50	134.26	0.5910	1150	012021 sus scrofa (pig). apom
SwissProt_38:GALU PSEAE -	96.00	145.44	0.5804	279	035633 pseudomonas aeruginosa
SwissProt_38:AMU2 XENIA -	94.00	131.99	1.04	875	012850 xenopus laevis (afric
SwissProt_38:YK3 CAEEL -	94.00	123.73	1.38	1895	014951 caenorhabditis elegans
SwissProt_38:PYR1 SCCHO -	92.00	118.84	2.19	2244	009794 schistosoma haematophyllum
SwissProt_38:MCS BOVIN -	90.00	128.88	2.40	563	038091 bos taurus (bovine).
SwissProt_38:NK1 YEAST -	90.00	122.61	2.80	1080	038662 saccharomyces cerevisiae
SwissProt_38:TUD DROME -	88.00	114.48	3.42	2515	025823 drosophila melanogaster
SwissProt_38:SSRS RAT -	88.00	129.84	3.30	363	030938 rattus norvegicus (rat)
SwissProt_38:AMU HUMAN -	87.00	118.71	5.13	974	019021 homo sapiens (human).
SwissProt_38:GCEI CHICK -	86.50	126.29	4.64	407	049707 gallus gallus (chicken)
SwissProt_38:PHV3 AVEA -	86.50	116.48	5.89	1128	006593 avena sativa (oat).
SwissProt_38:YAOE SCCHO -	86.50	113.41	6.05	1260	010093 schistosoma haematophyllum
SwissProt_38:SRB8 YEAST -	86.50	114.22	6.23	1427	025648 saccharomyces cerevisiae
SwissProt_38:DSG2 HUMA -	85.50	114.94	7.25	1117	014126 homo sapiens (human)
SwissProt_38:HRA FUGRU -	85.00	114.95	7.89	1025	042611 fuji rubripes (japan)
SwissProt_38:P11A MOUSE -	85.00	114.56	7.96	1068	042337 mus musculus (mouse)
SwissProt_38:ROD SPIDU -	85.00	112.22	8.43	1361	011704 spiraea oleracea (s
SwissProt_38:TOXA CLODI -	85.00	105.59	9.91	2710	016154 clostridium difficile
SwissProt_38:DHG CHLAE -	85.00	100.74	11.15	4485	039575 chlamydomonas reinhardtii
SwissProt_38:YHHT ECOLI -	84.50	124.51	6.80	349	037622 escherichia coli. hyf
SwissProt_38:NM5A PLTI -	84.50	122.95	7.06	410	012637 canis familiaris (dog)
SwissProt_38:CAQC CANPA -	84.50	120.06	7.58	554	034855 apis mellifera (honeybee)
SwissProt_38:ROB PLANA -	84.50	114.14	8.76	1024	021421 plasmodium falciparum
SwissProt_38:RMS5 EMENT -	84.00	122.14	7.84	410	023351 emericella nidulans (f
SwissProt_38:RMS5 NEUR -	84.00	121.77	7.91	426	023351 neurospora crassa. ml
SwissProt_38:AMU BOVIN -	84.00	113.83	9.60	972	010731 bos taurus (bovine). R
SwissProt_38:POLG ECL2T -	84.00	106.00	11.62	2193	006575 escherichia coli. typh
SwissProt_38:SYM SYNA -	83.00	118.00	10.28	532	005729 synecococcus sp. (str
SwissProt_38:AMU RAT -	83.00	112.16	11.85	976	014925 rattus norvegicus (rat).
SwissProt_38:KFL1 CHICK -	83.00	112.08	11.88	984	004861 gallus gallus (chicken)
SwissProt_38:IDPE DICLI -	82.50	119.96	9.44	237	022549 dictyostelium discoideum
SwissProt_38:PPP BACST -	82.50	119.16	10.88	433	017836 bacillus stearothermophilus
SwissProt_38:DVHC ONCMY -	82.50	117.49	11.33	515	015305 oncomycobacterium mykiss (f
SwissProt_38:TISA ECOLI -	82.50	116.20	11.69	589	019704 escherichia coli. typh
SwissProt_38:GRF DICLI +	82.50	114.43	12.21	708	036417 dictyostelium discoideum
SwissProt_38:Y502 CAEEL -	82.50	114.29	12.25	718	010128 caenorhabditis elegans
SwissProt_38:UJ52_HSVJ -	82.50	112.55	12.78	861	02468 herpes simplex virus

SwissProt_38:V33 BTV10 - 82.50 112.11 12.92 901 | P12435 bluetongue virus (s
SwissProt_38:PHV4 AVEA - 82.50 109.95 13.62 1128 | P06594 avena sativa (oat
SwissProt_38:YEM2 YEAST - 82.50 105.70 15.10 1753 | P32634 saccharomyces cer
SwissProt_38:P11A HUMAN - 82.00 109.66 14.93 1068 | P42336 homo sapiens (hum

seq_name: SwissProt_38:V33 CAEEL

seq_documentation_block:

ID	CUT1 CAEEL	STANDARD:	PRT:	424 AA.
AC	003755: 018693:			
DT	01-FEB-1994 (rel. 28, Created)			
DT	15-JUL-1998 (rel. 36, Last sequence update)			
DT	15-DEC-1998 (rel. 37, Last annotation update)			
DE	CUT1-1 PRECURSOR.			
GN	CUT-1 OR C47G2.1.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;			
OC	Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RX	MEDLINE: 91323673.			
RA	SEBASTIANO M., LASSANDRO F., BAZZICALUPO P.;			
RT	"cut-1 a Caenorhabditis elegans gene coding for a dauer-specific			
RL	noncollagenous component of the cuticle.";			
RL	Dev. Biol. 146:519-530(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RA	PALMER S.;			
RT	Submitted (Apr-1995) to the EMBL/Genbank/DBJ databases.			
CC	-1- FUNCTION: COMPONENT OF THE CUTICLES. IT FORMS A RIBBON			
CC	APPROXIMATELY 2 MICRONS WIDE RUNNING ALONG THE LATERAL LINES			
CC	UNDERNEATH THE ALAE. CONTRIBUTES TO THE FORMATION OF EXTRACELLULAR			
CC	ENVELOPES PROTECTING THE ORGANISM FROM THE ENVIRONMENT.			
CC	-1- TISSUE SPECIFICITY: LOCALIZED IN A NARROW BAND NOT EXPOSED			
CC	TO THE EXTERIOR, RUNNING UNDERNEATH THE ALAE.			
CC	-1- DEVELOPMENTAL STAGE: EXPRESSED DURING DAUER LARVA FORMATION.			
CC	-1- DOMAIN: THE SMALL REPEATS A-A-P-(AVI) ARE ALSO PRESENT IN MANY			
CC	PROTEINS CONSTITUTING THE PROTECTIVE ENVELOPE OF OTHER SPECIES.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: M55997; AAA27995.1; ALT_INIT.			
DR	EMBL: Z49125; CAN88934.1; -.			
DR	PIR: A49772; A49772.			
DR	HSSP: P04002; IATP.			
DR	WORMPEP: C47G2.1; CE02165.			
DR	Cuticle; Signal; Repeat.			
FT	SIGNAL	1	18	POTENTIAL.
FT	CHAIN	19	424	CUTICLIN 1.
FT	DOMAIN	302	323	4 X 4 AA REPEAT OF A-A-P-(AVI).
FT	REPEAT	302	305	1.
FT	REPEAT	307	311	2.
FT	REPEAT	312	315	3.
FT	REPEAT	320	323	4.
FT	CONFLICT	27	27	G -> P (IN REF. 1).
FT	CONFLICT	29	29	P -> G (IN REF. 1).
FT	CONFLICT	306	306	A -> G (IN REF. 1).
FT	CONFLICT	309	309	MISSING (IN REF. 1).
FT	CONFLICT	389	390	CL -> SS (IN REF. 1).
FT	CONFLICT	395	396	FA -> LP (IN REF. 1).
FT	CONFLICT	401	402	IG -> MR (IN REF. 1).
SQ	SEQUENCE	424 AA;	45168 MW;	08E03529 CRC32;

alignment_scores: Quality: 1420.50 Length: 416
 Ratio: 4.228 Gaps: 7
 Percent Similarity: 80.769 Percent Identity: 67.067

alignment_block:
 US-09-323-427-5/rev x CUT1_CAEBL ..

Align seg 1/1 to: CUT1_CAEBL from: 1 to: 424

```

1137 TGTACATCACTTATTCATGCTGTATTCG...ATCCGGTTCAGCATGC 1091
    ||| : : : : : ||| ||| ||| ||| ||| ||| ||| ||| |||
   8 CysLeuAlaIaLeuValIaLeuSerAlaIaIaIaIaIaIaIaIaIaIa 24
1090 TGTGCAAGTGAGCCAGAAATGATGTGGCAACATTCATTAACATCA 1041
    : : : : : : : : : : : : : : : : : : : : : : : : : :
   24 nValGIuGIuGIuProGIuValGIuGlySGIuProAsnSerIleThrVala 41
1040 ATTTTAATACAGTAAATGCATGCAAGACATGTTTATGTAAGCTTT 991
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   41 snPheAsnThrArgAsnProPheGIuGIuHisValIuValIuValIu 57
   57 TGTATGCAACAAGTGCCTGATGATGAGTGCAGTGCAGTGCCTG 941
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   58 TGTATGCAACAAGTGCCTGATGATGAGTGCAGTGCAGTGCCTG 941
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   940 AATTCACATTCATTCATGATGATGATGATGATGATGATGATGATG 891
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   74 yIleGIuLeuProPheAsnSerCysAsnThrAlaArgThrArgSerIe 91
   91 ATCCAGTGTGATTTTGTATACACAACATGTTGTCATTTGCTTACCA 841
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   91 snProIuSGIuValIaPheValSerThrThrValaIleSerPheHis 107
   840 TTTTGTGTTCCAAAGTGTGATGATGATGATGATGATGATGATGAT 791
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   108 GluPheValThrIuSValaIaPArgAlaIuArgIleGIuIncysPhe 124
   124 TGTuserAspIuSValaIaPArgAlaIuArgIleGIuIncysPhe 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   740 CTGCTTTTCAACTCAATTTGCCAGTATGCCAGTATGCCAGTATG 691
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   141 hrIaPheGIuThrGIuValIaIaPArgAlaIuArgIleGIuIncys 157
   690 TTTGATGTTGACCAACCGGTCAACAGTTCATTTGCTTACCTGTC 641
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   158 LeuAspGIuGIuProSerGIuGIuProIleGIuPheAlaThrIleGI 174
   174 GCCAGTTTATCATTAATGACATGCGATTCCTGAACCGTTCATTC 591
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   174 nGIuValIuThrHisIuSValaIaPArgAlaIuArgIleGIuIncys 191
   590 GCCGGTGTGTCATTCCTGCTTTGCTGATGATGATGATGATGATG 541
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   191 ySAlaValIaIaHisSerCysThrValaIaPArgAlaIuArgIle 207
   540 GAAATTCATTAATGCTGATGATGATGATGATGATGATGATGATG 491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   208 GIuIleLeuAsnGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGI 224
   490 TTTGGAATATCCAAAGATTTTATGCTGGCCAAAGATGCTGATTA 441
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   224 nIeGIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIu 241
   440 AATATGCGATGATGATGATGATGATGATGATGATGATGATGATG 391
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   241 ySTyAlaIaPArgSerGIuIuPheIuGIuIuIuIuIuIuIuIuI 257
   390 ATTAAGAACAATAGCGATGATGATGATGATGATGATGATGATGAT 341
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   258 IleIuSAspProGIuSerGIuCysAlaIaArgProIuThrCysSer 274

```

```

340 AGCATTCGAGATGTTTAAACAGTGT..... 313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   274 nGIuPheGIuAlaValIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 291
   312 .....GCCGCA 307
   291 IeAlaIaProGIuAlaGIuValaGIuGIuValaGIuAlaIaIaIa 307
   306 GCAAAACCTGCTGCAGCT..... 289
   308 AlaIaIaProValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 324
   288 .....GCCGCACTGCTTATTC...AAGAAAGATCT. 259
   324 IProArgAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 341
   258 ..GCAGACCGGAGAAATATGATGATGATGATGATGATGATGATG 211
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   341 heGIuGIuAsnGIuGIuIleLeuAspValaIaIaIaIaIaIaIa 357
   210 GAATATGAGATGATGATGATGATGATGATGATGATGATGATGAT 161
   358 AspIleMetGIuGIu.....AlaSerProSerAlaIaIaIaIaIa 372
   160 ACTTCTGCAACATTAATGACACACCTGTAATATCTGCTGCAAGT 111
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   372 AluValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 387
   110 GAATTCGATGATGATGATGATGATGATGATGATGATGATGATG 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   387 IuIleCysLeuThrProIleGIuPheAlaSerPheIuGIuIleGI 403
   60 GCATGATGCTGCTGCGCATTAATGATGATGATGATGATGATGAT 13
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   404 IleValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 419

seq_name: SwissProt_38:YOH3_CAEBL
seq_documentation_block:
ID YOH3_CAEBL STANDARD. PRT. 495 AA.
AC 009276.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 55.1 KD PROTEIN C43C3.3 IN CHROMOSOME X.
GN C43C3.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditi; Rhabditiida;
OC Rhabditiida; Rhabditiidae; Pelodierinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL NZ;
RA SULSTON J.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z47067; CAAB7330.1; -.
DR WORKREP: C43C3.3; CE01525.
KW Hypothetical protein; Transmembrane; ATP-binding.
FT TRANSMEM 11 31 POTENTIAL.
FT TRANSMEM 426 446 POTENTIAL.
FT NP_BIND 171 178 ATP (POTENTIAL).
FT SEQUENCE 495 AA; 55068 MW; 09C77468 CRC32;

```

alignment_scores:
 Quality: 108.00 Length: 340
 Ratio: 0.679 Gaps: 19
 Percent Similarity: 46.765 Percent Identity: 22.647

alignment_block:
 US-09-323-427-5/rev x YOH3_CAEEL ..

Align seg 1/1 to: YOH3_CAEEL from: 1 to: 495

```

1071 ATGTGATGTGGACCACTTCAATACAAATCAATTTAAT..... 1033
      ||||| ||||| ||||| ||||| |||||
83 ValAspCysIleAlaAspSerPheThrValIleuAsnLysSerAspPr 99
1032 .....ACAGTAATGCATTCGAAGACAGATGTTATG 1002
      ||||| ||||| ||||| ||||| |||||
99 OGluValMetArgMetIleSerAsnProLysSerGlnProValIleValTy 116
      ||||| ||||| ||||| ||||| |||||
1001 TGAAGGCTTTTATGATCAAGAAGTTGC.....CGTAATGATGAA 961
      ||||| ||||| ||||| ||||| |||||
116 AluArgLysIleLysThrArgHisProCysGlyThrSerMetLysAspGlu 132
      ||||| ||||| ||||| ||||| |||||
960 GGTGGACGTCAGTACGCCGAATTCCTTCATTTGATTC...TGCAA 914
      ||||| ||||| ||||| ||||| |||||
133 LysGlyLeuThrAsnPheAsnLeuThrIleProLysGlySerGluCysAs 149
      ||||| ||||| ||||| ||||| |||||
913 TGTGGGCGGTACACGATCTCTGATCCAGTGGTATTTTGTAAACAACA 864
      ||||| ||||| ||||| ||||| |||||
149 PAlaThrLeuThrAspLeu.....ProLysHisArgTyrAlaGluThrT 164
      ||||| ||||| ||||| ||||| |||||
863 CTGTGTCAATTCGTTTCAT...CCATTATTTGTACCAAGTTGATCGT 817
      ||||| ||||| ||||| ||||| |||||
164 hValValIleuGluAspAsnAlaAspLeuSerPheGlyLysThrThrArg 180
      ||||| ||||| ||||| ||||| |||||
816 GCATTCAGTACATGCTTTTACATGGAAGCTGATAAACAGTACTAGTC 767
      ||||| ||||| ||||| ||||| |||||
181 LeuAsnHisValPheCysLeuTyrThrArgAsnValLysThr..... 194
      ||||| ||||| ||||| ||||| |||||
766 ACAGATTGAGTATCTGAATCAACAACCTGCTTTCA.....A 729
      ||||| ||||| ||||| ||||| |||||
195 ....IleArgPheSerAspAlaSerAsnGlyHisGluValIleAlaSerT 210
      ||||| ||||| ||||| ||||| |||||
728 CTCGAATTTGCCGATGCGCAGTATGCCGTTATGAAATTTTG..... 688
      ||||| ||||| ||||| ||||| |||||
210 hrcLysGlyLysProLysPro.....LysValGluMetLeuPheArgSer 224
      ||||| ||||| ||||| ||||| |||||
687 ...GATGTGGACCAACCGGTCAACA.....GTTC 659
      ||||| ||||| ||||| ||||| |||||
225 ThrAspSerGlyLysThrLeuGlnAlaAlaArgGluAsnGluPheValG 241
      ||||| ||||| ||||| ||||| |||||
658 ATTGTCATCATTTGTCAGCCA.....GTTTATCATTAATGACA.... 619
      ||||| ||||| ||||| ||||| |||||
241 uPhePheIleAlaLeuSerProAspSerAlaTyrHisGlyIleSerProL 258
      ||||| ||||| ||||| ||||| |||||
618 .....TCGATTTGAAACCGTTGAT..... 598
      ||||| ||||| ||||| ||||| |||||
258 yscLysThrPheSerAspArgLysAspIleSerAlaProAspAlaLys 274
      ||||| ||||| ||||| ||||| |||||
597 .....ACTTTCGCGGGTTGTCATTCCTGCTTGTGATGATGATAA 554
      ||||| ||||| ||||| ||||| |||||
275 LysIleThrPhe.....ValGlnGlyGlyCysProValAsnGlyMetAs 289
      ||||| ||||| ||||| ||||| |||||
553 CGGTGATACGTGTGGAATCTCTAATGCTGATGATGCTCTTGATTAAT 504
      ||||| ||||| ||||| ||||| |||||
289 nAsp..... 290
      ||||| ||||| ||||| ||||| |||||
503 ATTTCGTAATAATTGGAATATCAACAGATTTAATGCTGGCCAAGAA 454
      ||||| ||||| ||||| ||||| |||||
291 ..IleIleAspProLeuAlaAsnValAsnAspGlnIleTyrPheSerLys 306
      ||||| ||||| ||||| ||||| |||||
453 GCTCACGATATACAAATATGCGGATCGATCAACAGCTTTTCTATCAATCCA 404
      ||||| ||||| ||||| ||||| |||||
307 PheArgThrPheArgPheGlyAsnGlnSerThrValPheValHisCysG 323
  
```

seq_documentation_block:
 ID YAY3_SCHPO STANDARD; PRT; 649 AA.
 AC 010211;
 DT 01-OCT-1996 (rel. 34, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 01-OCT-1996 (rel. 34, Last annotation update)
 DE HYPOTHETICAL 74.5 KD PROTEIN C4H3.03C IN CHROMOSOME 1.
 GN SPAC4H3.03C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Archaescomycetes;
 OC Schizosaccharomycetes; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972.
 RA MURPHY L., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
 RL Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: Z69380; CAA93342.1; -
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 265 285 POTENTIAL.
 FT TRANSMEM 564 584 POTENTIAL.
 FT TRANSMEM 626 646 POTENTIAL.
 SO SEQUENCE 649 AA; 74488 MW; F04F8763 CRC32;

alignment_scores:
 Quality: 97.50 Length: 403
 Ratio: 0.594 Gaps: 23
 Percent Similarity: 43.672 Percent Identity: 22.333

alignment_block:
 US-09-323-427-5/rev x YAY3_SCHPO ..

Align seg 1/1 to: YAY3_SCHPO from: 1 to: 649

```

1089 GTCGAAGTACGCCAGAAATTTGATGTGGACCACTTCAATACATCA 1040
      ||||| ||||| ||||| ||||| |||||
36 LeuAspLysSerValGluMetMetCysTrpPro.....As 47
1039 TTTTAATACAGTAATGCATTC.....GAAGGAC 1011
  
```



```

CC -1- SUBUNIT: INTERMOLECULAR DISULFIDE BONDS COULD HELP MAINTAIN A
CC MULTIMERIC LOCIN STRUCTURE.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: SUBMAXILLARY MUCOSAE.
CC -1- DOMAIN: CONTAINS TANDEMLY REPEATED, IDENTICAL SEQUENCES OF 81
CC RESIDUES.
CC -1- PM: EXTENSIVELY O-LINKED GLYCOSYLATED ON SER AND THR RESIDUES OF
CC THE REPEAT UNITS. HIGHEST GLYCOSYLATION APPEARS TO OCCUR ON SER
CC RESIDUES WHICH HAVE GLY AT POSITIONS AT +2 OR -2 FROM THE
CC GLYCOSYLATION SITE OR, WHERE GLY IS THE PENULTIMATE RESIDUE, THE
CC PRESENCE OF PROLINE (USUALLY AT POSITION +3 OR -3) APPEARS TO ALSO
CC ENHANCE GLYCOSYLATION.
CC -1- SIMILARITY: CONTAINS 1 WFEC DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M61883; AAA30998.1; -.
DR EMBL: M21174; AAA30990.1; -.
DR PIR: A40009; A40009.
DR PROSITE: PS01185; CTCK_1; 1.
DR PROSITE: PS01225; CTCK_2; 1.
DR PROSITE: PS01208; WFEC; 1.
DR PFAM: PF00007; Cys_knot; 1.
DR Glycoprotein; Repeat; 1.
FT NON_TER 1
FT DOMAIN 1 368 81 AA TANDEM REPEATS.
FT REPEAT <1 44 1 (FRAGMENT).
FT REPEAT 45 125 2.
FT REPEAT 126 206 3.
FT REPEAT 207 287 4.
FT REPEAT 288 368 5.
FT REPEAT 369 391 6 (INCOMPLETE).
FT DOMAIN 929 995 WFEC.
FT DOMAIN 1062 1146 CTCK.
FT DISULFID 1062 1109 BY SIMILARITY.
FT DISULFID 1076 1123 BY SIMILARITY.
FT DISULFID 1085 1139 BY SIMILARITY.
FT DISULFID 1089 1141 BY SIMILARITY.
FT DISULFID ? 1145 BY SIMILARITY.
FT CARBOHYD 46 46
FT CARBOHYD 50 50
FT CARBOHYD 51 51
FT CARBOHYD 57 57
FT CARBOHYD 58 58
FT CARBOHYD 61 61
FT CARBOHYD 66 66
FT CARBOHYD 67 67
FT CARBOHYD 73 73
FT CARBOHYD 74 74
FT CARBOHYD 76 76
FT CARBOHYD 77 77
FT CARBOHYD 81 81
FT CARBOHYD 83 83
FT CARBOHYD 87 87
FT CARBOHYD 91 91
FT CARBOHYD 93 93
FT CARBOHYD 94 94
FT CARBOHYD 96 96
FT CARBOHYD 98 98
FT CARBOHYD 101 101
FT CARBOHYD 103 103
FT CARBOHYD 104 104
FT CARBOHYD 106 106
FT CARBOHYD 107 107
FT CARBOHYD 108 108
FT CARBOHYD 110 110

```

```

FT CARBOHYD 114 114
FT CARBOHYD 117 117
FT CARBOHYD 123 123
FT CARBOHYD 124 124
FT CARBOHYD 418 418
FT CARBOHYD 547 547
FT CARBOHYD 917 917
FT CARBOHYD 985 985
FT CARBOHYD 1002 1002
FT CARBOHYD 1068 1068
SQ SEQUENCE 1150 AA; 109615 MW; F7C55CCE CRC32;

```

```

alignment_scores:
  Quality: 98.00      Length: 221
  Ratio: 1.089      Gaps: 12
  Percent Similarity: 40.724      Percent Identity: 23.982

```

alignment_block:

US-09-323-427-5/rev x APMU_PIG ..

```

Align seg 1/1 to: APMU_PIG from: 1 to: 1150

780 AAACAGTTAGTGCACAGATTGAGTATCTGAA...ATCACAACTGCTTT 734
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
892 LysgluiletherhralaserProlysalserProgluThrAlaGl 908
733 TCAACTGCA.....A 723
908 yAlaThrGluaspGlnGluasnGluasnLysThrGlyCysProAlaProl 925
722 TTGTCCCGATGCCAGTATGCCGTTATGAATTTGGATGGTGACCAACC 673
    :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
925 euProProProProValCysHis.....GlyProLeu 935
672 GGTCAACAGTTCAATTTGCTATCATTTGTCACACAGTTATGATAATG 623
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
936 GlyGluGluLysSerProGlyAspValTyrThrAlaAsnGlyHisLysCys 952
622 GACATGC...GATTCTGAACCGTTGAT.....ACTTCTGGCGG 585
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
952 sThrCysThrGluAlaLysThrValaLysCysLysProLysGluCysProS 969
584 TTGTCCATTCCTGCTTGTTCGATGAT.....GCTTAC 553
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
969 euProProThrCysLysThrGlyLysArgLeuLeuLysPheLysAlaAsn 985
552 GGTGATACCTGTGGAATTTCTAAATGCTGATGATGCTCTTGATAAATA 503
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
986 AspThrCysCysGluLeu.....GlyHisCysGluLysArgThrCys 999
502 TTGTCTAAATTAATTTGCAATATCCACACAGATTAAATGGCTGGCCAGAG 453
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
999 sLeuPheAsnAsnThrAspTyr..... 1006
452 CTCACGATATACAAATATGCGGATGCATCACAGCTTTCTATCAATGCCAG 403
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1007 .....Glu 1007
402 ATCAGTATTACCATTAAGACCAATAGCGAATGTTTCGACACCAATG 353
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1008 ValGlySerSerPheAspAspProAsnAsnProCysValThrThrSerCys 1024
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
352 TTCAGAACCAACAGATTCGAGCTGTAA.....AACA 319
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1024 s...GlnAsnThrGlyPheThrAlaValaValaGlnAsnCysProLysGlnT 1040
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
318 GGTGTGTCGCCG.....AGCAAAACTGCTGC... 293
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1040 hTrTyrCysAlaGluGluAspArgValTyrAspSerLysGlnCysCysTyr 1056
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
292 AGCTGGCAACTTCGTTTACTACAGAAAAGATCTGCAGAACCGAGATA 243
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```


SQ SEQUENCE 1895 AA; 208393 MW; FEA6D6FA CRC32;

alignment_scores:
 Quality: 94.00 Length: 209
 Ratio: 0.940 Gaps: 8
 Percent Similarity: 47.847 Percent Identity: 21.053

alignment_block:
 US-09-323-427-5/rev x YLK3_CAEEL ..

Align seg 1/1 to: YLK3_CAEEL from: 1 to: 1895

```

916 CAATGTGGCGCTACACGATCTCTGAATCAGCGTGTATTTTGAACA 867
||||| : : : : : |||||
1149 GInCysIleAspAsnSerValCysMetAsnGlnMeCysThrCysAsnAs 1165
866 CAAC.....TGTTCATTTTCGTTTCATCATTTATT 835
||||| : : : : : |||||
1165 nAsnTyrArgLeuValTyrGlyTyrCysValProIleThrSerIleC 1182
1182 YSgIglInThrIleuVal..... 1189
784 TGATAAACAGTTAGTCACAGATGAGTATCTGAATCACAAGCTGTT 735
||||| : : : : : |||||
1190 .....AsnAsnGlnCysValIleuSerIleValGlyIleThrCysI 1204
734 TTCAACTCAAAATTTGCCGATGCCAGTATGCGCTTAAATTTTGAT 685
||||| : : : : : |||||
1204 eAlAsnGlnGlnCysValGlyIleValMetCys.....AsnSerGlyT 1219
684 GGTGACCAACCGGTCAACCAAGTCAATTTGCTATCATGTGACCGCT 635
1219 hrcYsGlnCysThrAsnGlyAlaThrAlaMetIleGlyIleThrCys 1235
634 TTATCATCAATGACATGCGATTCGAAACCGTGTATCTTCGCGCG 585
||||| : : : : : |||||
1236 SerSerSerSerCysAsnSerAsnGlnVal.....SerI 1248
584 TTGTTCATTCCTCGCTTTTGCATGATGTGACGCGATCTGCAAAAT 535
||||| : : : : : |||||
1248 IeAsnGlyMetCysTyr.....AsnThrValAlaIleVal 1258
534 CTAAATGCGATGATGATGCTCTTGTATAAATTTGCTAAATATTGGA 485
1259 .....GlyGlySerCysSerPheSerGlnGlnCysLeuAsnAlaVal 1273
484 ATATCCAACAGATTTAATGCTGCGCAAGACCTCAGCATATCAATAAT 435
||||| : : : : : |||||
1273 IcyThrAsnAsnIleCysValSerThrPheCysSerValSerCysSer 1290
434 CGGATCGATCACAGCTTTTCTATCAATGCCAGATAGTATTCACAT 385
||||| : : : : : |||||
1290 hrasnGlnValCysIleSerAsnGlnCysTyrAsnTyrValSerIleGly 1306
384 GAA.....CCAAATAGGAATGTGTTGCG 362
||||| : : : : : |||||
1307 SerGlnCysValGlySerGlnGlnCysLeuSerAsnSerGlnCysIle 1323
1323 rSerIleCysGlnCysProGlnGly 1331
seq_name: SwissProt_38:PYRL_SCHPO
seq_documentation_block:
ID PYRL_SCHPO STANDARD; PRT; 2244 AA.
AC 009794;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)

```

```

DE URAL PROTEIN [INCLUDES: GLUTAMINE-DEPENDENT CARBAMOYL-PHOSPHATE
DE SYNTHASE (EC 6.3.5.5); ASPARTATE CARBAMOYLTRANSFERASE (EC 2.1.3.2)].
GN URAL OR SPAC267.06c.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
RN [1]
RP SEQUENCE OF 22-2244 FROM N.A.
RC STRAIN-972;
RX MEDLINE; 96020160.
RA LOULIER M., JACQUET L., NEDEVA T., LACROUTE F., POTIER S.,
RA SOUCIET J.-L.;
RT "As in Saccharomyces cerevisiae, aspartate transcarbamoylase is
RT assembled on a multifunctional protein including a dihydroorotase-like
RT cryptic domain in Schizosaccharomyces pombe."
RL Curr. Genet. 28:138-149(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA BADCOCK K., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS PROTEIN IS A "FUSION" PROTEIN ENCODING THREE
CC ENZYMAIC ACTIVITIES OF THE PYRIMIDINE PATHWAY (GATASE, CPASAE,
CC AND ATCASE) (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 2 ATP + GLUTAMINE + CO(2) + H(2)O = 2 ADP +
CC ORTHOPHOSPHATE + GLUTAMATE + CARBAMOYL PHOSPHATE.
CC -1- CATALYTIC ACTIVITY: CARBAMOYL-PHOSPHATE + ASPARTATE =
CC ORTHOPHOSPHATE + N-CARBAMOYLASPARTATE.
CC -1- PATHWAY: FIRST AND SECOND STEP IN PYRIMIDINE BIOSYNTHESIS.
CC -1- MISCELLANEOUS: GATASE (GLUTAMINE AMIDOTRANSFERASE) AND CPASAE
CC (CARBAMOYL PHOSPHATE SYNTHASE) FORM TOGETHER THE
CC GLUTAMINE-DEPENDENT CPASAE (GD-CPASAE) (EC 6.3.5.5)
CC -1- MISCELLANEOUS: IN EUKARYOTES EC 6.3.5.5 IS SYNTHESIZED BY TWO
CC PATHWAY-SPECIFIC (ARGININE AND PYRIMIDINE) UNDER SEPARATE CONTROL.
CC -1- SIMILARITY: THE CPASAE DOMAIN IS SIMILAR TO OTHER CPASAEs.
CC -1- SIMILARITY: THE PROTEIN CONTAINS A DEFECTIVE DHASE DOMAIN.
CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-1 GLUTAMINE
CC AMIDOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X81841; CA57433.1; -.
DR EMBL; Z54328; CA91130.1; -.
DR HSSP; P00968; IJDB.
DR PROSITE; PS00097; CARBAMOYLTRANSFERASE, 1.
DR PROSITE; PS00442; GATASE_TYPE_1; 1.
DR PROSITE; PS00866; CPASAE_1; 2.
DR PROSITE; PS00867; CPASAE_2; 2.
DR PFAM; PF00117; GATase; 1.
DR PFAM; PF00185; OTCase; 1.
DR PFAM; PF00289; CPsase_L_chain; 2.
DR PFAM; PF00744; Dihydroorotase; 1.
DR PFAM; PF00988; CPsase_sm_chain; 1.
KW Pyrimidine biosynthesis; Ligase; Transferase; Multifunctional enzyme.
KW Pyrimidine biosynthesis; Ligase; Transferase; Multifunctional enzyme.
FT DOMAIN 1 437
FT DOMAIN 438 477
FT DOMAIN 478 1514
FT DOMAIN 1515 1524
FT DOMAIN 1525 1853
FT DOMAIN 1854 1935
FT
FT

```


DR EMBL; M36192; AAA30657.1; -.
 DR HSSP; P00214; 1FPC.
 DR PROSITE; PS01208; VMFC; 1.
 DR PROSITE; PS01185; CTCK_1; 1.
 DR PROSITE; PS01225; CTCK_2; 1.
 DR PFM; PF00007; Cys_knoc; 1.
 KW Repeat; Glycoprotein.
 FT DOMAIN 61 158 3 x 11 AA REPEATS.
 FT REPEAT 61 71 1.
 FT REPEAT 112 122 2.
 FT REPEAT 148 158 3.
 FT DOMAIN 338 404 VMFC.
 FT DOMAIN 471 555 CTCK.
 FT SIMILAR 1 282 TO PORCINE APOMUCLIN.
 FT DISULFID 471 518 BY SIMILARITY.
 FT DISULFID 485 532 BY SIMILARITY.
 FT DISULFID 494 548 BY SIMILARITY.
 FT DISULFID 498 550 BY SIMILARITY.
 FT DISULFID ? 554 BY SIMILARITY.
 FT CARBOHYD 28 28 POTENTIAL.
 FT CARBOHYD 394 394 POTENTIAL.
 FT CARBOHYD 477 477 POTENTIAL.
 SQ SEQUENCE 563 AA; 58913 MW; B8YD4189 CRC32;

alignment_scores:
 Quality: 90.00 Length: 345
 Ratio: 0.612 Gaps: 18
 Percent Similarity: 42.609 Percent Identity: 20.290

alignment_block:

US-09-323-427-5/rev x MUCS_BOVIN ..

Align seg 1/1 to: MUCS_BOVIN from: 1 to: 563

```

1101 GTTGCAATGCTGTCGAGTGCAGTGCAGCAATGATGTCGACCACTTC 1052
      ::::::::::::::: |||
192  ILeGIthrhGlyllethrhGlythrhGlyser.....GlythrhTsr 205
      ::::::::::::::: |||
1051 AATAACATCAATTTTAAATACAGT...AATGCATTGAGAGACAGTGT 1005
      ::::::::::::::: |||
205  rSerPrroGlyglyPheasnaIaGluIaIarhtrhrPheIyGluIaIa 222
      ::::::::::::::: |||
1004 ANGTGAAGGCTTTATGATCAAGAAGTGGCCGTAATGATGTAAGTGCA 955
      ::::::::::::::: |||
222  rghtrhrgIuIthrhArgIleLeuSerGlythrhtrhArgIyArGserGly 238
      ::::::::::::::: |||
954  CGTCAAGTTGCCGGAATTCACCTTCATTGATTCATGCAATGTCGCCG 905
      ::::::::::::::: |||
239  ThtrhVal.....IlePro...GluSerSerasnThrhGlyTh 250
      ::::::::::::::: |||
904  TACACGAGCTCTGATATCCAGTGGTATTTTGTAAACAACAGTGTGTCA 855
      ::::::::::::::: |||
250  rSerThrhGlyValGlyArgIln.....ThrSerThAlaVal 263
      ::::::::::::::: |||
854  TTTTCGTTTCATCCATTATTTGTTACC..... 829
      ::::::::::::::: |||
263  aIserGlyArGValThrhGlyValSerGluSerSerPrroGlyThSer 279
      ::::::::::::::: |||
829  ..... 829
280  LysGluIaIaSerGluThrhtrhGlyProGlyIleSerThrhGlySe 296
828  .....AAAGTTGATCGTATGATGATGATGATGATGATGATGATG 795
296  rThrSerLysSerAsnArgIleThrhtrhSerSerArgIlePro.....T 311
794  ACATGAAGCTGTATAAACAAGTAGTGCACAGATGAGATGATGTAATC 745
      ::::::::::::::: |||
311  yPrroGluThrhtrhValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 326
744  ACATGCGCTTTTCAAACTCAATATGTCGCCGATGCCAGTATAGCCGTTATGA 695
      ::::::::::::::: |||

```

```

327  LysThrhGlyCysThrhtrhSerIeuProProProIaIaCysTyr..... 341
694  AATTTGGATGATGGACCAACCGTCACAGCTTCATATTGCTATGATG 645
      ||||| ||||| :::: |||
342  .....GlyProIeuGlyGluIyLysSerPrroGlyIaIaIaIaIaIa 354
644  GTCAGCAGCTTTATCAATAATGACATGC...GATTCTGAACCGTTGAT 598
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
354  rPthraIaIaIaCysHISLysCysThrhCysThrasPalaGluThraIaIa 370
597  .....ACTTCGCGGGGTTCGCCATTCCTGCTTTGTCATGATG 557
      ||| ||| :::: |||
371  CysLysIeuLysGluCysProSerProProThCysLysProGluGlu.. 386
556  TAACGGTATGATCTGGAATTTCTAATGCTGATGATGATGCTGATGATA 507
      ||| ||| :::: ||| |||||
387  .....ArgIeuValLysPheLysAsnAsnAsnAsnAsnAsnAsnAsn 401
506  AATAT.....TTGCTAATAATATTTGGAATATCAACA 475
      ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
401  LatYrCysGluProArGThrhCysLeuPheAsnAsnAsnAsnAsnAsn 415
474  GATTTATGCTGTCGCCAAGACCTCACGTATACAAATATGCGGATGCATC 425
415  .....AACAGTGTGCGCC..... 415
424  ACAGCTTTCTCATATGCGACATCAATATACCAATTAAGACCAATA 375
      ::::::::::::::: |||
416  .....GluValGlyIaIaSerPheIaIaIaIaIaIaIaIaIaIaIaIa 426
374  GCGAATGCTGTGACCAACATGTTCAAGAACACACAGATTCGAGCTGT 325
      ::::::::::::::: |||
426  snPrroCysIleSerTyrSerCysHISAsn...ThrhGlyPheValaIaIa 441
324  AA.....AACAGTGTGCGCC..... 308
442  ValGlaIaIaPcysProLysGluIaIaIaIaIaIaIaIaIaIaIaIaIa 458
307  ....ASCAAAACCTGCTGCGAGTGCAGTGCACAACTGCTTACTCAAGAA 262
      ||| ||||| :::: |||
458  rAsPserThrhLysCysCys.....TyrThrhCysLysPrroT 470
261  TCTGCAAGACCGGAGATATCATGATGTAAGCAAC 227
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
470  yrcYsaIaIaSerSerSerValaIaIaIaIaIaIaIaIaIaIaIaIaIa 481
seq_name: Swissprot_38:NRK1_YEAST
seq_documentation_block:
ID NRK1_YEAST STANDARD; PRT; 1080 AA.
AC P38692;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE SERINE/THROMBIN-PROTEIN KINASE NRK1 (EC 2.7.1.-) (N-RICH KINASE 1).
GN NRK1 OR YHR102W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC-5;
RA FUKAMI Y.;
RL Submitted (MAY-1994) to the EMBL/Genbank/DBJ databases.
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE; 94378003.
RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DOVER J.,
RA DU Z., FAVELLO A., FULTON L., GATUNG S., GEISL C., KIRSTEN J.,
RA KUCABA T., HILLIER L., JIER M., JOHNSTON L., LANGSTON Y.,
RA LATREILLE P., LOUIS E.J., MACRI C., MARDIS E., MENDEZ S., MOUSER L.,
RA NHAN M., RIFKIN L., RILES L., ST PETER H., TREVASKIS E., VAUGHAN K.,

```

RA VIGNATI D., WILCOX L., WOHLMAN P., WATERSTON R., WILSON R.,
 RA VAUDIN M.;
 RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
 VIII.";
 RT Science 265:2077-2082(1994).
 CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
 CC PROTEIN KINASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D29980; BAA06250.1; -;
 DR EMBL; U00059; AAB68860.1; -;
 DR PIR; S48944; S48944.
 DR HSSP; P24941; 1A01.
 DR SGD; L0001276; NRK1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP_1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST_1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM_1.
 DR PFAM; PF00069; PKinase; 1.
 KM Transferase; Serine/threonine-protein kinase; ATP-binding.
 FT DOMAIN 23 276 PROTEIN KINASE.
 FT NP_BIND 29 37 ATP (BY SIMILARITY).
 FT BINDING 52 52 ATP (BY SIMILARITY).
 FT ACT_SITE 144 144 BY SIMILARITY.
 SQ SEQUENCE 1080 AA; 117061 MW; E1668B6F CRC32;

alignment_scores:
 Quality: 90.00 Length: 209
 Ratio: 0.833 Gaps: 6
 Percent Similarity: 51.675 Percent Identity: 21.531

alignment_block:

US-09-323-427-5/rev x NRK1_YEAST ..

Align seg 1/1 to: NRK1_YEAST from: 1 to: 1080

```

722 CAACATCAATGTCGCCGATGCCGATGCCGTTATGAAATTTGGATGG 683
      :::::::::::  ::  ::  :::::::::::  ||
200 LysValAspIleTyrSerLeuGlyIleThrTyrGluIleAlaThrCl 216
      :::::::::::  ::  ::  :::::::::::  ||
682 TGGACCA.....ACCGGTCACACGATTCATTTGCTA 651
      :::::::::::  ::  ::  :::::::::::  |
216 YasnProProTyrCysAspValGluAlaLeuArgAlaMetGlnLeuIleI 233
      :::::::::::  ::  ::  :::::::::::  |
650 TCAATGTCACGCCAGTTTATCATTAATGACATGCGATTCGAAACCGTT 601
      :::::::::::  ::  ::  :::::::::::  ||
233 IleYserIysProProArgLeuGlnAspArgSerTyrSerThrSerLeu 249
      :::::::::::  ::  ::  :::::::::::  ||
600 GATACCTTCGCGCGGTTCATTCCTCTTGTCGATGATGTAACGG 551
      ||  ||:::  ||:::  ||:::  ||:::  ||:::  ||:::  ||:::  ||:::  ||
250 LysGluPheIleAlaLeu.....CysLeuAspGlnAspProLys.. 262
      :::::::::::  ::  ::  :::::::::::  ||
550 TGAATCTGTGAATTTCTAATGCTGATGATGTGCTTGTATTAATATT 501
      ||  ||:::  ||:::  ||:::  ||:::  ||:::  ||:::  ||:::  ||
263 .....GluArgLeuSerAlaAspAspLeuLeuLysSerLysPheI 276
      :::::::::::  ::  ::  :::::::::::  ||
500 TGCCTAAATATTTGGAATTCACACAGATTTAATGGCTGGCCAGACAGCT 451
      ::  ::  ::  ||:::  ||:::  ||:::  ||:::  ||:::  ||:::  ||
276 LeArgAlaIhisLysAlaThrProThrSerIleLeuLysGluLeuLieser 292
      ::  ::  ::  ||:::  ||:::  ||:::  ||:::  ||:::  ||:::  ||
450 CACGATATACAAATATGCGGATGCACAGCTTTTCTATCAATGCCAGAT 401
      ::  ::  ::  ||:::  ||:::  ||:::  ||:::  ||:::  ||:::  ||
293 ArgTyrLeuLeuPheArgAspLysAsnLysTyrLysIleGluI 309
      ||:::  ||:::  ||:::  ||:::  ||:::  ||:::  ||:::  ||:::  ||
400 CAGTATTCACATTAAGAACAATAGCGAATGTGTTCACCAACATATGT 351
      ||:::  ||:::  ||:::  ||:::  ||:::  ||:::  ||:::  ||:::  ||

```

```

309 ySerIleProGluAsnGluProSerLysProSerGluAlaProLysProS 326
350 CAGAACCCACAGAGATTGGAGCTGTAAACAGGNGMGCCACCAAAA 301
      ||:::  ||:::  ||:::  ||:::  ||:::  ||:::  ||:::  ||:::  ||
326 ergLn.....AsnGlyGlyGlyAspGluAlaGln 335
      ::  ::  ::  ||:::  ||:::  ||:::  ||:::  ||:::  ||
300 CCTGTGCGAGTCGCCAAGCTTGTACTCAAGAAAGATTCGCAAGAC 251
      ::  ::  ::  ||:::  ||:::  ||:::  ||:::  ||:::  ||
336 LysSerIleAlaSerAsnAspAsnGluIleLysArgValAsn..... 349
      ::  ::  ::  ||:::  ||:::  ||:::  ||:::  ||:::  ||
250 GGAAGATATCATTTGATGTACGACATGATATCAACACCTTGAATTAGCG 201
      ||:::  ||:::  ||:::  ||:::  ||:::  ||:::  ||:::  ||
350 GluGlyAspValGluMetLysTyrAspPheAspSerLeuSerSerA 366
      ||:::  ||:::  ||:::  ||:::  ||:::  ||:::  ||:::  ||
200 AT.....GATATCAAGCTTTGCCATGTATTA 172
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
366 sPTyrIleIleGluAsnAsnIleAsnLeuAspAlaLeuAlaGluAspAsn 382
      ::  ::  ::  ||:::  ||:::  ||:::  ||:::  ||:::  ||
171 CGTCACCGTCGACATTCTGCACACATTAAT 145
      ::  ::  ::  ||:::  ||:::  ||:::  ||:::  ||:::  ||
383 AsnGluTrpAlaThrAlaGlnHisAsp 391
      ::  ::  ::  ||:::  ||:::  ||:::  ||:::  ||:::  ||

```

seq_name: SwissProt_38:TUD_DROME

seq_documentation_block:

```

ID TUD_DROME STANDARD; PRT; 2515 AA.
AC P25823;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE MATERNAL TUDOR PROTEIN.
TUD.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP MEDLINE: 92038995.
RX COLUMBESKI G.S., BARDSLEY A., TAX F., BOSWELL R.E.;
RT "Tudor, a posteriori-group gene of Drosophila melanogaster, encodes a
RT novel protein and an mRNA localized during mid-oogenesis.";
RL Genes Dev. 5:2060-2070(1991).
CC -1- FUNCTION: REQUIRED DURING OOGENESIS FOR THE FORMATION OF
CC PRIMORDIAL GERM CELLS AND FOR NORMAL ABDOMINAL SEGMENTATION.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT THE LIFE CYCLE.
CC -1- MISCELLANEOUS: THE TUD mRNA ACCUMULATES WITHIN THE POSTERIOR
CC REGION OF THE DEVELOPING OOCYTE DURING THE EARLY TO MIDDLE STAGES
CC OF OOOGENESIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62420; CAA44286.1; -;
DR PIR; A41519; A41519.
DR FLYBASE; FBgn0003891; tud.
DR PFAM; PF00567; TUDOR; 10.
KW Developmental protein
SQ SEQUENCE 2515 AA; 285236 MW; 1C381DB3 CRC32;

```

alignment_scores:
 Quality: 90.00 Length: 367
 Ratio: 0.552 Gaps: 16
 Percent Similarity: 44.414 Percent Identity: 19.074

alignment_block:

US-09-323-427-5/rev x TUD_DROME ..

Align seg 1/1 to: TUD_DROME from: 1 to: 2515

```

1116 TCCTATTGATTCGGGTGACATGATGTCGAGGAGCCAGCAATATGA 1067
1117 ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1904 SerLeuGlnLeuPro...AspAlaIyrIleSerTrpSerProGlnLagi 1919
1066 ATGTGGACCACTTCATACACA.....ATCAATTTTAATACAC 1029
1117 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1919 uAlaIyrPheAlaGlnLeuThrGlyGlnGlyLeuValPheThrThrc 1936
1028 GTCATGTCATTCAGAGA.....CATGTTATGTGAAGGCTCTTATGAT 965
1117 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1936 LndeuLeuIysProGlyGlnAspHisValThrIleAspLeuLeuAsp 1952
984 CAGAGAGTTGGCCGTATGATGTAAGGAGCAGTCAGTTGCCGAATTTC 935
1953 .....GlyGlnAsnIleIleAspArgLe 1960
934 ACTTCGATTTGATTCGACATGTCGGCGTACAGCATCTCTGATCCAC 885
1117 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1960 uIeuPProLeu.....CysGlnIarGlySgluProLysGluAlaSerLysg 1975
884 GTGGTATTTTGTACACACACTGTGTCTATTTCGTTTCATCCATTATT 835
1117 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1975 LndeuLeuAlaValThrThrLysAlaIleIle..... 1985
834 GTTACCAAGTTGATGTCGATTCGATACATGCTTTACATGGAAGC 785
1117 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1986 ...ThnHisValGlnAsnThrSerArgIleTyrLeuGlnPheSerGly 2001
784 TGAT..... 781
2001 sAspSerLeuMetAspIleIleCysGlnLysLeuAsnGlySerLysLeuG 2018
780 ....AAACAGTTAGTCACAGATTGAG..... 757
2018 LndeuLysThrGlnLysAlaAlaValAspAspMetCysValValGlnPhe 2034
756 .....GATCTGAATACCAACTGC 737
2035 AlaAspAspLeuGlnPheTyrArgSerArgIleLeuGlnValLeuGlnAs 2051
736 TTTTCAACTCAATGTCGCCAGTCCAGTATGCCGTTATGAATTTTGG 687
2051 pAspGlnTyrLysValIle.....LeuIleLe 2060
686 ATGTGTGACCAACC.....GCTACACAGTTTCAA 658
2060 sPtyrGlyAsnThrThrValValAspLysLeuTyrGlnLeuProGlnGlu 2076
657 TTTGATCATGTTGTCAGCCAGTTTATCATTAATGACATCGATTCTGA 608
2077 PheThrLeuIle...LysProValAlaGln.....IleCysSerMetG 2090
607 A.....ACCGTTGATACTTCTTCT 591
2090 uProSerAlaIlePheGlnLysAsnLysAlaLeuThrLeuThrPhePhe 2107
590 GCGCGGTTCCATTCCTGCTTTGTGCATGATGCTAACGGTACTGCG 541
2107 sPAlaLeuLeuAspSerCys..... 2113
540 GAAATCTTAATGCTGATGATGCTGCTTGAATAATATTGCTAATAATA 491
2114 .....LysGlyValValAlaValGlnPheValAsnLysSe 2125
490 TTTGGAATTCACACAGATTTAATGCTGCGCCAGAGAGCTCACGATATAC 441
2125 rAlaSerProProValValArgLeuThrThrLysAspLysArgSerLeu 2142
440 AATATGCGATGCATCACAGCTTTTTCATCAATGCCAGATCAGTATTACC 391

```

```

2142 yslIeTyrGlnHisLeuGlnLysLeuValGlnAlaGlnLeuLysLeuIle 2158
390 ATTAAG...GAACCAATATAGCGCAATGTGTCGACCAACAATGTTCAGACCC 344
2159 GlnLysArgAsnGlnAsnSerGlnCysIleIleSerTyrGlyAsnSerTr 2175
343 ACAAGAGTTTCGGAGCTGTTAAACAGAGTGTGGCCGACGACAAAACCTG 294
2175 oLysSerPheTyrValGlnMetLysHisAsnSerAlaAspLeuAspLeuI 2192
293 CAGCTGCCCACTTCCTTACTTCACAGAAAAGATCTGCAGAACCGAGANT 244
2192 LeuLysThrLeuGlnSerLeuLysGlnLysLeu.....LysLys 2206
243 ATCATGATGATAGCACTGATATACACACCCCTTGAATTAAGCATATTA 194
2207 LeuLleAspProThrThrAsnSerAsnGlyValCysTyrSerGlnLys 2223
193 T 193
2223 p 2223

seq_name: SwissProt_38:SSR5_RAT
seq_documentation_block:
ID SSR5_RAT STANDARD; PRT; 363 AA.
AC P30938;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE SOMATOSTATIN RECEPTOR TYPE 5 (SSSR).
GN SSTR5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PIUITARY;
RX MEDLINE: 93125499.
RA O'CARROLL A. M., LOIAIT S. J., KONIG M., MAHAN L. C.;
RT "Molecular cloning and expression of a pituitary somatostatin
RT receptor with preferential affinity for somatostatin-28.";
RL Mol. Pharmacol. 42:939-946(1992).
RN [2]
RP REVISIONS TO C-TERMINUS.
RC TISSUE=PIUITARY;
RX MEDLINE: 94195267.
RA PANETTA R., GREENWOOD M. T., WARSZYNSKA A., DEMCHISHYN L. L., DAY R.,
RA NIZNIK H. B., SRIKANT C. B., PATEL Y. C.;
RT "Molecular cloning, functional characterization, and chromosomal
RT localization of a human somatostatin receptor (somatostatin receptor
RT type 5) with preferential affinity for somatostatin-28.";
RL Mol. Pharmacol. 45:417-427(1994).
CC -!- FUNCTION: RECEPTOR FOR SOMATOSTATIN-28. THE ACTIVITY OF THIS
CC RECEPTOR IS MEDIATED BY G PROTEINS WHICH INHIBIT ADENYLYL CYCLASE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: PROMINENT IN THE PITUITARY AND SMALL
CC INTESTINE. LOW LEVELS IN ISLETS AND SPLEEN. NOT DETECTED IN
CC KIDNEY, PANCREAS, CEREBELLUM, OR CORTEX.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: L04535; AAA17029.1; -.
CC EMBL: U01152; AAC09011.1; -.
CC EMBL: X74828; CAA52825.1; -.
CC GCRDB: GCR_0514; -.

```



```

927 TTTGATTCATGCAATGTTGGCGCTACAGATCT..... 895
    ||| ||||| ||||| |||||
89 PheIleHisLeuAsnValAlaSerThrArgSerSerProLeuProIleLe 105
894 .....CTGAATCCACGTC 882
105 uGlyTrpAlaAsnArgAspArgValTrpLysAsnMetIleAsnLysGlu 122
    :||| :|
881 GATTTTGTGAACACACAGCTGTCATTCGTTTCATCATCATTAATTT... 835
    :||| :|
122 LuThrTyValArgAspLysLeuTyMetGlnArgHisProLeuLeuGln 138
834 .....GTTACCAAGT 824
139 ProLysMetArgThrIleLeuLeuAspTrpLeuMetGluValLysGluVal 155
823 TGAATCGCATATGAGATCAATGCTTTATACAGAACTGAT..... 781
    | :||| :|
155 LTrpLysLeuTyArg...GluThrPheTyIleuAlaGlnAspPhe 171
780 .....AAACAGTTATGACACAG 763
171 SPArgPheMetAlaThrGlnGlnAsnValLysThrLeuLeuGlnLeu 187
762 ATTGAGTATCTGAAATCAACAACTGCTTTCAA...ACTCAATTTGCC 716
    ||| :||| :|
188 IleGlyIleSerSerLeuPheIleAlaLysLeuGluGluLeuTyTrp 204
715 GATGCAGATGCCGTTATGAA...ATTTGGATGCTGACCAACCGGTC 669
    :||| :|
204 oPolLysLeuHisGlnPheAlaTyValThrAspGlyAlaCysThr.Glu 220
668 AACCACTTCATTTGCTATCATGTCGTACAGCAAGTTATCATATGAGACA 619
    :||| :|
221 AspGlnIleLeuSerMetGluLeuIleIleMetLysAlaLeuAsnTrpAs 237
618 TGGCATTTGCAAAACCGTTGATA.....CTTTCGCCGGGTTT 581
    :||| :|
237 n.....LeuAsnProLeuThrValAlaSerTrpLeuAsnIleTyMetG 252
580 CCAATTCCTGCTTTGCGATGATGTAACGCTGATCTG..... 541
    :||| :|
252 InValAlaTyIleuAsnGluLeuTyGluValLeuLeuProGlnTyPro 268
540 .....GAATTTCTAATGCTGATGATGTC 515
269 GlnGlnIlePheValAlaIleAlaGluLeu.....AspLeuGlyVal 283
514 TCTTGATTAATATTTGCTAATATTTGGAATATCCAAACAGATTTAATG 465
    :||| :|
283 IleuAsp.....IleGlyLysLeuGluTyTrpIleTyGluValLeu 297
464 CTGGCCAGAGAGCTCAAGTATCAAAATATGCGGATGATCAGACGTT... 418
    :||| :|
297 IaIleSerIleLeuTyHisPheSerSerSerGluLeuMetGlnLysVal 313
417 .....TTCATCAATGCGCAGATCATATTACATTAAGAACAATAG 374
314 SerGlyTyGluIleTyProLysGluLe.....G 322
373 CGAATGCTGTCGACCAAGTATGTCAGAACCAAGGATTCGAGCTGTTA 324
    :||| :|
322 uGlyLysValLysTrpMetValProPheAlaMet.....AlaIle 336
323 AAACAGAGTGGCCGACGCAAAACCTGCTGACGCTGCCAAGCTTGTTA 274
    :||| :|
336 rGluValAlaGlySerSer.....LysLeu 343
273 CTCAGAAAAAGATCTGCAGAACCGGAATATCATGATGATAGCACTGA 224
    :||| :|
344 LysHisPheArgGlyIleAlaProGlnAspLeuHisAsnIleGlnThrH 360

```

```

223 TATCAACACCCCTTGAATTAAGCATGATATATCAACCT 187
    ||||| :|||
360 sIleAsnSerLeuAspLeuLeuAsnLysAlaGlnAla 372
seq_name: SwissProt_38:PHY3_AVEsa
seq_documentation_block:
ID PHY3_AVEsa STANDARD: PRT: 1128 AA.
AC P06593:
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PHYTOCHROME A TYPE 3 (AP3).
GN PHYA3 OR PHY3.
OS Avena sativa (Oat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC Poaceae; Avena.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 86093660.
RA HERSHEY H.P., BARKER R.F., IDLER K.B., LISSEMORE J.L., QUAIL P.H.;
RT "Analysis of cloned cDNA and genomic sequences for phytochrome:
RT complete amino acid sequences for two gene products expressed in
RT etiolated Avena."
RL Nucleic Acids Res. 13:8543-8559(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98185837.
RA HERSHEY H.P., BARKER R.F., IDLER K.B., MURRAY M.G., QUAIL P.H.;
RT "Nucleotide sequence and characterization of a gene encoding the
RT phytochrome polypeptide from Avena."
RL Gene 61:339-348(1987).
RN [3]
RP SEQUENCE OF 1-12.
RA GRIMM R., KELLERMAN J., SCHAEFER W., RUEDIGER W.;
RT "The amino-terminal structure of oat phytochrome."
RL FEBS Lett. 234:497-499(1988).
CC -!- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT
CC ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS
CC MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT
CC ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCOVERSION OF PR IN
CC PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS
CC RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE
CC RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR
CC GENES, INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-
CC BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,
CC PROTOCHLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS
CC THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.
CC -!- SUBUNIT: HOMODIMER.
CC -!- PM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.
CC -!- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X03242; CAA26999.1; -.
DR EMBL: M18822; AAA76820.1; -.
DR PIR: A29631; A29631.
DR PIR: S00096; S00096.
DR PIR: S03032; S03032.
DR MENDEL: 1305; AVEsa; PHYA.1.
DR PROSITE: PS00245; PHYTOCHROME_1; 1.
DR PROSITE: PS0046; PHYTOCHROME_2; 1.
DR PFAM: PF00360; phytochrome; 1.
DR PFAM: PF00512; signal; 1.
DR PFAM: PF00989; PAS; 2.
DR PFAM: PF01590; GAF; 1.

```

KW	Transcription regulation; Photoreceptor; Phytochrome; Chromophore;
KM	Multigene family.
FT	INT_MET 0
FT	BINDING 321 321 CHROMOPHORE.
FT	CONFLICT 278 278 L -> F (IN REF. 2).
SD	SEQUENCE 1128 AA; 124861 MW; 49DA285 CRC22;

[illegible]

```
alignment_block:
```

Align seg 1/1 to: PHY3_AVEA from: 1 to: 1128

```

691 TTTGGATTGGGGAGCCAAACGGTCACACAGTTCATATTTCTTATCTTGGTC 642
692 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
771 PheGlyTrp.....CysSerGluTrpAs 778
641 AGCCAGCTTTATCATATAA.....TGCAGATGGCATTTGAAACCGTT 601
642 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
778 AlaAlaMetThr..LysLeuThrGlyTrpAsnArgAsp...GluValLeu 793
600 GATACATTTCTGGCGCGTTCATTCCTGCTTTGTTCGATATGCTAAACG 551
601 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
794 AspLysMet.....LeuLeuGlyGluValPhe.....AspSerSerAs 806
550 TGAATACGAGGAATCTTAATAGCTGAGGATGTCCTTATGATTAATTT 501
551 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
806 AlaSerCysProLeuLysAsnArgAspAlaPheValSerLeuCysValL 823
500 TCGTAAATATTTGGCAATATCCACACAGATTTAATGCTGCGCCAGAAAGCT 451
501 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
823 euLleAsnSer.....AlaLeuAlaGlyGluGluThr 833
450 CAC.....GTATACAAATATGCGGATCGATCGATACAGCTTTCTATCATG 407
451 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
834 GluLysAlaProPheGlyLysPheAspArgSerGlyLysTyrIleGly 850
406 CCAGATCAGTATTATTCATTAAAGAACCAATAGCGAA..... 370
407 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
850 sLeuLeuSerAlaAsnArgLysGluAsnGluGlyLeuIleThrGly 867
369 .....TGTTGGACACACATTTTCAGAACCCACAGAAGTTGCGAGCGTT 325
370 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
867 alphecys..PheIleHis...ValAlaSerHisGluLeuGlnHisAlaLe 882
324 A.....AAACAGGTGTCGCCGACCAAAACCTGTCGATCGTGGCGCACT 281
325 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
882 uGlnValGlnGlnAlaSerGlnGlnThrSerLeuLysArgIleuLysAla 899
280 TCGTTTACTCAAGAAAGATCTGCAGAACCGGAGAAATATCATTTGATGAC 231
281 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
899 heSerTyrMetArgHisAlaIleAsnAsnProLeuSerGlyMetLeuTyr 915
230 GA.....ACTATATTCAC.....AC 215
916 SerArgLysAlaLeuLysAsnThrAspLeuAsnGluGlnMetLysG 932
917 CTTTGAATTTAGCGATGAT.....AATCAACCTTTGGCAG 180
918 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
932 nIleIleHisValGlyAspAsnCysHisIleGlnIleAsnLysIleLeu...A 948
179 TTGATTTTACGTCACCGTCGACCTTTCGAACATAATGCA..... 142
179 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
948 lAspSerLysAspGlnAspSerIleThrGluLysSerSerCysLeuAspLeu 964
141 .....CAACCTTATACTCTGTCGATGACA.. 115
665 GluMetAlaGluPheLeuLeuGlnAspValValValAlaAlaValSerG 981

```

```

114 .....AATGCAATC.....TGCAATGTAC 96
981 nValleuIleIeThmCysGlnGlyLysGlyIleArgIleSerCysAsnLeuP 998
95 CATTGGCTTCATCAATGTTATG.....GGTTTA 67
|| ::::| | | | | | | | | | | | | | | | | |
998 ro.....GluArgPheMetLysGlnSerValTyrGlyAspGlyAla 1011
66 AGCATTGCATTGATGCTGCGCTCATTTATACCATTTGCTTAAATTTG 17
::: || | | | | | | | | | | | | | | | | | |
1012 ArgGlnGlnGlnIleLeuSerAspPheLeuPheIleSerValLysPheSe 1028
16 TCCTA 13
|| | |
1028 rPro 1029

```


THIS PAGE BLANK (USPTO)

OM of: US-09-323-427-5 to: SPFREMBL_12:★ out_format : pfs
Date: Apr 16, 2000 4:56 AM
About: Results were produced by the GenCore software, version 4.5
Copyright (c) 1993-2000 CompuGen Ltd.

```

Command line parameters:
-MODEL=framer-n2p.model
-o=/cgn2.1/USPO9323427/runat_14042000.170514_19920/app_query.fasta.1.1
-DB=SPRMBL_12_QEWT-fastcan -SUFFY=bscktrans.rpt
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOCL=0.000
-LOOPEXT=0.500 -QGAOP=4.500 -QGAPEXT=0.050 -XGAOP=10.000
-XGAPEXT=0.500 -FGAOP=6.000 -FGAPEXT=7.000 -YGAOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blomsium62 -TRANS=human40.cdi -LIST=45 -DOTALIGN=200
-THR_SCORE=pct -ALIGN=15 -MODE=LOCAL -OUTEXT=pfs -NORM=ext
-MINLEN=0 -MAXLEN=1000000 -USER=USO9323427 -NCPU=6 -ICPU=3
-NO_XLIFY -WAIT -THREADS=1

```

```
Search information block:
Query: US-09-323-427-5
Query length: 1161
Database: SPTREMBL_12: *
Database sequences: 225878
Database length: 6933412
Search time (sec): 253.700000
```

Sequence	Strid	Orig	Zscore	EScore	Len	Documentation
sp_invertebrate:0193177	-	163.30	2830.20	2.8e-150	385	! 0931071 ascaris lumbricoides (human)
sp_invertebrate:0939707	-	1382.50	2363.51	2.7e-124	389	! 019707 caenorhabditis elegans
sp_invertebrate:0184719	-	1313.50	2243.77	1.2e-117	433	! 018479 meloidogyne artiella (human)
sp_invertebrate:0939707	-	725.50	1232.91	2.8e-61	364	! 093971 caenorhabditis elegans
sp_invertebrate:0939707	-	695.00	1185.43	2.2e-58	206	! 093971 wuchereria bancrofti (human)
sp_invertebrate:0936775	-	662.00	1132.88	2.9e-55	127	! 0936775 brugia malayi (human)
sp_invertebrate:0218080	-	654.00	1112.68	1.7e-54	290	! 021808 caenorhabditis elegans
sp_invertebrate:0215450	-	602.50	1007.49	1.8e-49	480	! 021540 caenorhabditis elegans
sp_invertebrate:0190503	-	595.00	1005.71	9.4e-49	554	! 019053 caenorhabditis elegans
sp_invertebrate:0935352	-	578.50	981.09	3.4e-47	315	! 093532 caenorhabditis elegans
sp_invertebrate:0221624	-	546.00	924.25	4.1e-44	384	! 022164 caenorhabditis elegans
sp_invertebrate:0939707	-	511.00	859.04	1.1e-40	610	! 093971 caenorhabditis elegans
sp_invertebrate:0939367	-	477.50	804.24	1.7e-37	440	! 093967 caenorhabditis elegans
sp_invertebrate:0230977	-	476.00	802.88	2.4e-37	387	! 023097 caenorhabditis elegans
sp_invertebrate:0226680	-	453.50	760.05	3.6e-35	609	! 022680 caenorhabditis elegans
sp_invertebrate:018213	-	414.00	696.16	2.1e-31	742	! 018213 caenorhabditis elegans
sp_invertebrate:009586	-	409.50	682.49	6.1e-31	747	! 009586 caenorhabditis elegans
sp_invertebrate:0215733	-	401.50	669.67	3.5e-30	670	! 021573 caenorhabditis elegans
sp_invertebrate:0967776	-	361.00	620.43	2.0e-26	66	! 096776 brugia pahangi (human)
sp_invertebrate:0191304	-	357.50	597.70	5.5e-26	437	! 019304 caenorhabditis elegans
sp_invertebrate:0618186	-	339.00	571.71	3.0e-24	225	! 061816 caenorhabditis elegans
sp_invertebrate:022627	-	293.00	488.21	8.3e-20	366	! 022627 caenorhabditis elegans
sp_invertebrate:045890	-	290.50	484.45	1.4e-19	344	! 045890 caenorhabditis elegans
sp_invertebrate:016304	-	158.00	252.77	7.7e-07	514	! 016304 caenorhabditis elegans
sp_invertebrate:0233587	-	141.00	206.53	4.2e-05	3507	! 023587 caenorhabditis elegans
sp_invertebrate:0181177	-	137.50	209.54	8.0e-05	1262	! 018117 caenorhabditis elegans
sp_invertebrate:0227274	-	132.50	206.59	0.0002	665	! 022724 caenorhabditis elegans
sp_invertebrate:024328	-	126.50	195.85	0.0008	697	! 024328 dirosophila melanogaster (human)
sp_invertebrate:0019494	-	120.00	181.37	0.0037	1011	! 001944 caenorhabditis elegans
sp_invertebrate:020167	-	116.00	177.11	0.0086	751	! 020167 caenorhabditis elegans
sp_invertebrate:0182828	-	110.50	165.64	0.0299	942	! 018282 caenorhabditis elegans
sp_invertebrate:0972288	-	105.00	148.12	0.1138	2340	! 097288 plasmodium falciparum (human)
sp_invertebrate:020550	-	103.50	154.72	0.1380	828	! 020550 caenorhabditis elegans
sp_mammal:062672	-	101.00	144.66	0.2615	1589	! 062672 bos taurus (bovine)
sp_vertebrate:0919597	-	100.00	147.62	0.3038	935	! 091597 xenopus laevis (african)
sp_human:005707	-	99.50	147.68	0.3346	843	! 005707 homo sapiens (human)
sp_mammal:018758	-	98.00	120.72	0.6735	13288	! 018758 sus scrofa domestica (human)
sp_mammal:077509	-	95.00	135.95	0.9598	1323	! 077509 bos taurus (bovine)
sp_bacteria:0929EG6	-	94.00	139.55	1.10	724	! 0929EG6 carobacterium divergens
sp_bacteria:051229	-	93.50	140.21	1.21	610	! 051229 borrelia burgdorferi (human)
sp_invertebrate:0937239	-	93.00	122.87	1.73	3933	! 097239 plasmodium falciparum (human)

```

seq_name: sp_invertebrate:Q93115
ID Q93115 PRELIMINARY; PRT. 385 AA.
AC Q93115;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DR 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DI 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE CUT-1-LIKE CUTICLIN PROTEIN PRECURSOR.
GN ASCUT-1.
OS Ascaris lumbricoides (common roundworm).
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilia; Ascaridida;
OC Ascaridoidea; Ascarididae; Ascaris.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97390131.
RA TIMINOUNI M., BAZZICALUPO P.;
RT "cut-1-like genes of Ascaris lumbricoides.";
RL Gene 193:81-87(1997).
DR EMBL, 073005; AAB6646.1; -.
KW Signal.
FT SIGNAL. 1 16 POTENTIAL.
FT CHAIN 17 385 CUT-1-LIKE CUTICLIN PROTEIN.
SQ SEQUENCE 385 AA; 42408 MW; BDBE1169 CRC32;

alignment_scores:
    Quality: 1653.50      Length: 388
    Ratio: 4.593          Gaps: 4
    Percent Similarity: 92.784      Percent Identity: 80.155

alignment_block:
US-09-323-427-5/rev x Q93115 ..

Align seg 1/1 to: Q93115 from: 1 to: 385

1158 ATGATGTCGGTTTATGCTTCTGTACTACACGTAATGTCCTATATC 1109
||| ||| ::::::::::| ::::::::::|:::|:::|:::|
1 MeCysArgAlaValSerPhe...LeuAlaLeuPheGlyLeuAlaAlaI 16
1108 GATTCGCGTTCGAATGTCGTCGACGTGACGCCAAGTAATGATCGAC 1059
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||
16 alleProValAspAsnGlyValGluGluProGluIleGluCysGlyP 33
1058 CAATCTCAATACAAATCAATTTTAATACAGTAATGATCGAAGACAT 1009
|||||:|||||:|||||:|||||:|||||:|||||:|||||
33 roThrSerIleThrValAsnPheAsnThrArgAsnProPheGluGlyHis 49
1008 GTTTATCTGAAAGGCTCTTTATGATCAAGAAGGTGCCGTAATATGAAG 959
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||
50 AlaIyrValIysGlyLeuIyrAspGlnGluGlyCysArgSerAspGluI 66
958 TGGACGCGCAAGTGGCCGAATTCACCTTCATTCGATTCGATCGAATGTTG 909
|||||:|||||:|||||:|||||:|||||:|||||:|||||
66 yGlyAlaGlnValAlaGlyIleSerLeuProPheAspSerCysAsnValA 83
908 CGCGTACACGATCTCGAATCGACGTGTAATTTTAAACAACAAGCTGTT 859
|||||:|||||:|||||:|||||:|||||:|||||:|||||
83 laArgThrArgSerLeuAsnProArgGlyIlePheValIhrThrVal 99
858 GTCAATTCGTTTCATCCATTATTGTTACCAAGTTCGTCGATATCG 809
|||||:|||||:|||||:|||||:|||||:|||||:|||||
100 ValIleSerPheHisProLeuPheIleThrIlyValAspArgIaIyrAr 116
808 AGTACATAGCTTTTACATGGAAGCTGATAAACAGTGAATGACAGATG 759
|||||:|||||:|||||:|||||:|||||:|||||:|||||
116 yValGlnCysPheIyrMetGlnAlaAspIyrThrValSerThrGlnIleG 133

```

```

758 AGGTATCTGAAATCACACTGCTTTTCAAACTCAATGTCCCGATGCCA 709
133 lvalSerGIuIleThrAlaPheGlnThrGlnIleValProMetPro 149
708 GATGCCGCTTAAATTTTGGATGGTGACCAACCGGTCACACCATTCA 659
150 valCysArgTGIuIleLeuAspGIyGIyProThGIyGlnProIleGI 166
658 ATTGTGATCATTTGTCAGCAGCTTATCATTAATGACATGCGATTCTG 609
166 nheAlaThrIleGIyGlnGlnIleValThIstIyStrpThrCysAspSerg 183
608 AAACCGTGATACTTCTCGCGGGTTCATTCCTGCTTTTGTGCATGAT 559
183 luthrAlaSpThrPheCysAlaIleValIleHisSerCysPheValaPaSp 199
558 GGTAACGGTGATCTGTGGAAATTCATTAATGCTGATGATGTCCTTTTA 509
200 GIySerGIyAspThrIleGlnIleLeuAsnGIuGIyGIyCysAlaLeuAs 216
508 TAAATTTTGTAAATTAATTTGGAATATCCACAGATTAATGCGTGCGC 459
216 plYstYrIleLeuAsnAsnLeuGIuTyrProThrAspLeuMetAlaGIyG 233
458 AAGAACTCAGCTATACAAATATGCGGATCGATCACAGCTTTTCTATCAA 409
233 lGlnIuAlaHisValIyIyIstYrAlaAspArgSerGIuLeuPheTyrGI 249
408 TGGCAGATGATTTACCATTTAAGAACCAATAGGAAATGTCTTGACG 359
250 CysGIuIleThrIleThrIleGIyGIyGIyProAsnSerGIyCysProArgPr 266
358 ACAATGTTTCAGAACCAAGAGATTCGAGCTGTAAACAGCGTGCGC. 310
266 ohrCysSerGIuProGIyGIyPheGIyAlaValaArgProGIyGIySerI 283
309 ..GCAGCAAAACCTGCTGCAGCTGCGCAACTTCGTTTACTCAAGAAAGA 262
283 lAlaAProIySIySgInaArgArgCysGIuLeuArgLeuIleIySIySer 299
261 TCTGCAGAACCGAGAAATCATGTGATGACAGCTGATATCAACACCT 212
300 GIyGIyAspTyrAspAsnThrLeuAspValaArgThrAspPheSerIle 316
211 TCAAAATTCAGATGATATCAAGCTTGGCAGTTCGATTTACGTCACCGTG 162
316 uAspIleSerAspArgAspGIuAlaIleuProMetAspLeuArgHisArgA 333
161 CACTTCTGCAACAT..AATGCAACAACCTGTATATCTGCTGCAGTACAA 115
333 lA.....ArgHisAlaArgGIyGIyGlnGlnIleValIleuSerProAlaAsn 347
114 AATGCAATTCGATGTCACCATTTGGCTTCTCAATGTTTATGGGTTAAG 65
348 GIuGIyIleCysMetSerProPheGIyPheSerIlePheMetGIyLeuAl 364
64 CATTGCGATGATTCGCGCTGATTTATCACTTCGTTTAAATTCGCTG 15
364 aValAlaIleuAlaIleAlaIleValIleValIleValIleSerPheIySIyLeuArgP 381
14 CAATTCAGAAAGCA 1
381 roGIuGIyIleVala 385

```

seq_name: sp_invertebrate:Q19707

seq_documentation_block:

```

ID Q19707 PRELIMINARY; PRT; 389 AA.
AC Q19707;
DT 01-NOV-1996 (TReMBLrel. 01. Created)
DT 01-NOV-1996 (TReMBLrel. 01. Last sequence update)
DE 01-JAN-1999 (TReMBLrel. 09. Last annotation update)
DE F22B5.3 PROTEIN.

```

```

GN F22B5.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA SIMS M.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
RP [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BEKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIEH M., JOHNSTON L.,
RA JONES M., KERSHAM J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans".
RL Nature 368:32-38(1994).
DR EMBL; Z50044; CAA90355.1; -.
DR SEQUENCE 389 AA; 42983 MW; 56B9543C CRC32;

```

alignment_scores:

Quality: 1382.50 Length: 385
Ratio: 4.177 Gaps: 8
Percent Similarity: 85.974 Percent Identity: 69.610

alignment_block:

US-09-323-427-5/rev x Q19707 ..

Align seg 1/1 to: Q19707 from: 1 to: 389

```

1128 CTTATTCATTTGCTTTATTCGATTCGCGTTGACAAATGCGTGAAGTGA 1079
13 LeuValAlaISerValSerAlaIleProValaAspAsnValaGIuGIyGI 29
1078 GCCAGAAATGGAATGTGAGCAACACTTCAATTAACAATCAATTTTATATAC 1029
29 uProGIuValaGIuCysGIyProThrSerIleThrValaAsnPheAsnThra 46
1028 GTAATTCATTCGAGAGACATGTTTATGTGAAGCTTTTATATATCAAGA 979
46 rGAsnAlaPheGIuGIyHisValaIyValaIySIyGIyLeuPheAspGIuGI 62
978 GGTGCGCGTAATGATGAAGGTGAGCGTCAAGTTCGCGAATTCACCTTCC 929
63 GIuCysArgAsnAspGIuGIyGIyArgGIuValaIleAlaGIyIleGIuLeuPr 79
928 ATTTGATTCATGCAATGTTGCCGCTACACAGATCTGTAATCCACGTTGTA 879
79 oPheAspThrCysAsnValaIlaArgThrArgSerLeuAsnProIySgIy 96
878 TTTTCTATACACACTGTTGTCATTTGCTTTCATCCATTAATTATTGTTACC 829
96 AlPheValIleThrThrThrValaIleValSerPheHisProGIuIlePheValaIThr 112
828 AAGTTGATCGTGCATATGAGTATACATGCTTTTACATGAGTGAAGTGA 779
113 lYsValaAspArgAlaIyIyArgValaGIuCysPheIyIyMetGIuAlaAspIy 129
778 AACAGTTAGTGACAGATTTGAGCTATCTGAATCACAACACTGCTTTTCAAA 729
129 sThrValSerThrGlnIleGIuValSerAspLeuThrThrAlaPheGIuT 146
728 CTCGAATTTGCCGATGCCAGATATGCCGTTATGAATTTTGGATGGTGA 679
146 hrcIuValaIleProMetProIleCysIyStYrGIuIleLeuAsnGIyGI 162

```

```

678 CCAACGGCTACACCACTTCATTTGCTATCATGTGTGACGACGATTATGCA 629
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
163 ProthGlyGluProValGlnPheIatHrIleGlyGlnGlnValTyrH 179
628 TAAATGACATGGCATTTGGAACCGTTGATCTTCTGGCGGTTGTC 579
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
179 slvstPThrCysAspsrclutHrValAspThrPheCysAlaValValH 196
578 ATTCCGCTTTGCGATGATGTAACGGTGATCTGGAAATTCGAAT 529
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
196 lssercysThrValAspAspGlyAsnGlyAspThrValGlnIleuAsp 212
528 GCTGATGATGTGCTCTGATTAATATTGCTAAATATTGGATATGCC 479
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
213 GluAsnGlyCysAlaLeuAspLysPheIleuAsnAsnLeuGlyTyrPr 229
478 AACGATTTAAATGGCTGGCCCAAGACCTCAGTATCAAAATATGCGATC 429
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
229 oHrAspLeuMetAlaGlyGlnGlnAlaHisValTyrLysTyrAlaAsp 246
428 GATCAGACGCTTTCTATCAATGCCAGATCATTTACATTAAGAACCA 379
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
246 rgserrlneuphetyrGlnCysGlnIleSerIleThrValLysGluPro 262
378 AATAGCGAATGTGTGACACCAATGTTCAGAACCAAGATTCGAGC 329
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
263 AsnGlnGlnCysAlaArgProThrCysSerGluProGlnGlyPheGlyAl 279
328 TGTAAACAGAGTGTGGCCGACCAAAACCTGCTGACGTCGCGAACTTC 279
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
279 aValLysGlnAlaAsnGlnThrIleGln.....PhePhe 291
278 GTTACTGCAAGAAAGATCTGCGAGAA...CCGGAATATCATGTATGTA 222
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
291 rGValIleuLysLysArgSerAlaProValMetGluAsnIleuAspVal 307
231 CGAAGTATATCAACACCTTGAATATAGCATGATGAATCAAGCTTGGC 182
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
308 ArgAlaGlnLeuThrThrIleuGlnValLeuGlnGlyAsn.....LeuPr 322
181 AGTGAATTTACGTCACCGTCGACTTCTG.....CAACATAAATGAC 141
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
322 oSerSerLeuThrGlnAlaGlnAlaLeuValAlaSerArgGluIleGlyG 339
140 AACCTGTAATACTGCTGCGATGCAAAATGGAATTCGATGCACCATTT 91
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
339 lAsp.....SerPheArgGlnGlnLeuLysIleSerSerPhe 351
90 GCGTCTCA.....ATGTTATGGGTTTAAGCATTCGATTGATTCG 50
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
352 HisIleSerValValThrValAlaPheLeuGlyLeuThrVal.....PheVa 366
49 TGGCGTCATTAATACC.....ATTTCGTTTAAATTTGTCGTCGAATC 9
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
366 lAlaIlePheIleThrTyrMetIleValSerArgMetMetValProSerA 383
8 AGAAG 4
:|||||
383 sPlvs 384

```

seq_name: sp_invertebrate:018479

```

seq_documentation_block:
ID 018479 PRELIMINARY: PRT: 433 AA.
AC 018479:
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DE 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE CUTCUTIN-1.
CN MTCUT-1.
OS Meloidogyna artlellia.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Diplogasteria; Tylenchida;

```

```

OC Tylenchina; Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97218031.
RA DE GIORGI C., DE LUCA F., DI VITO M., LAMBERTI F.;
RT "Modulation of expression at the level of splicing of cut-1 RNA in the
RT Infective second-stage juvenile of the plant parasitic nematode
RT Meloidogyna artlellia.";
RL Mol. Gen. Genet. 253:589-598(1997).
DR EMBL; X96677; CA65452.1; -.
KW CUTCUTIN.
SQ SEQUENCE 433 AA; 47582 MW; FDE26884 CRC32;

alignment_scores:
Quality: 1313.50 Length: 418
Ratio: 4.042 Gaps: 7
Percent Similarity: 77.751 Percent Identity: 63.636

alignment_block:
US-09-323-427-5/rev x 018479 ..

Align seg 1/1 to: 018479 from: 1 to: 433

1158 ATGATTCGCTTATGCTTCTGCTACTACACTTATTCATTCGCT..... 1114
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MetArgLysIleuLeuPheAlaIleGlyValPheValAlaLeuAsnAlaI 17
1113 .....TATTGCAATTCGGTTGACAAATGGTGTGGAAGTGACCCAG 1074
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17 ePheThrValArgAlaIleProValAspAsnGlyValGlnGluProG 34
1073 AAATGAAATGGTGGACCACTTCATTAACAATCAATTTTAATACAGTAT 1024
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
34 lIleGlnGlyGlnProThrSerIleThrValAsnAsnThrArgAsn 50
1023 GCATTCGAAGACATGTTTATGGAAGGCTTTATGATCAAGAAGTTG 974
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 ProPheGlnGlyHisValTyrValLysGlyLeuPheAspGlnAlaGlyCy 67
973 CCGTAATGATGAAGGTGACAGTCGAAGTCCCGGAATTTTCACCTTCAT 924
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
67 sArgSerAspGlnHisGlyArgGlnValAlaGlyIleGlnLeuProPheA 84
923 ATTTCATGCAATGTTGGCGCTACACGATCTCGAATCCAGCTGATTTT 874
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
84 sPserCysAsnValAlaArgThr...AspAlaGlnProLysGlyValPhe 99
873 GTAAACAACAAGTGTGCAATTCGTTTACATCCATTTATGTTGTAACCAAGT 824
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
100 ValSerThrThrValValIleSerPheHisProGlnPheValThrLysVa 116
823 TGAATCGCATATCGAGTACAAATGCTTTTACATGGAAGCTGAATAACAG 774
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
116 lAspArgAlaTyrArgValGlnCysPheTyrMetGluAlaAspLysThrV 133
773 TTAGTCCACAGATGAGGTATCTGAAATCACAACGCTTTTCAACTGCA 724
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
133 alSerAlaGlnLeuGlnValSerGlnIleThrThrGlnIlePheGlnThrGln 149
723 ATTGTCCGATGCGCACTATGCGCTTATGAATTTTGGATGGTGGACCAAC 674
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
150 ValValProMetProValCysLysTyrGlnIleLeuGlnGlyAlaIle 166
673 CGGTCAACCAAGTTCAATTTGCTATCATGTTGGTCAGCCAGTTTATCATAAAT 624
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
166 uGlnGlnProIleGlnPheAlaThrIleAlaGlnGlnValTyrThrSerG 183
623 GGACATGCAATTCGAAACCGTTGATCTTCTGCGCGGTGTCCATTCG 574
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
183 lYThrCysAspSerGlnThrIleAspThrPheCysAlaValAlaHisSer 199
573 TGGTTTGTGATGATGTAACGGTGATCTGGAATAATCTGAATGCTGTA 524

```



```

215 p1ysglmetllelleleullelysgluaspserphecysmetvalvalh 232
578 ATTCCGCTTTGTCGATGATGTAACGGGTGATCTGTGAATTCGAAT 529
232 lsfThCysseValaspaspGlyArgGlyGluThSerPheleulleasp 248
528 GCTGATGATGCTGCTTGATTAATTAATTTGCTAAATTAATTTGGAATATCC 479
249 SerAnnglyCysSerileaspIysPheleuleSerAnsnleugluluyrrr 265
478 AACAGATTATAGCTGGCCAGAAAGCTCAGCTATACAAATATGCGGATC 429
265 ogIysnleuleleuA1agIyngluA1ah1sValTyrlYsPheA1aAspA 282
428 GATCAGAGCTTTTCTATCAATGACAGATCATATACATTAAGAAGCA 379
282 rGAspAlaleuPhePheGlnGlnleSerileThrVallysgluPro 298
378 AATAGCGAATGTGTCGACACACATGTCGAAACACAGATTCGAGC 329
299 aspGlnGluCysValarProIleCysgluaspValGluGlyGlyAl 315
328 TGTAAACAGGT 316
315 aproValValGly 319

```

seq_name: sp_invertebrate:09xy09

seq_documentation_block:

```

ID 09xy09 PRELIMINARY; PRT; 206 AA.
AC 09xy09;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE CUTICULIN-1 (FRAGMENT).
OS Wuchereria bancrofti.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;
OC Filarioidea; Onchocercidae; Wuchereria.
RN [1]
RP SEQUENCE FROM N.A.
RA RAMZY R., HELMY H., ADELY M., CUFTIS K., WEIL G.;
RT "Wuchereria bancrofti U3 cuticulin-1 cDNA partial sequence.";
RL Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF125580; AAD28743.1; -.
FT NON_TER 1
SQ SEQUENCE 206 AA; 22602 MW; 4DEE20D0 CRC32;

```

alignment_scores:

```

Quality: 695.00 Length: 198
Ratio: 4.064 Gaps: 4
Percent Similarity: 86.364 Percent Identity: 69.192

```

alignment_block:

us-09-323-427-5/rev x 09xy09 ..

Align seg 1/1 to: 09xy09 from: 1 to: 206

```

606 ACCGTTGATCTTTCTGCGGGTGTCCATTCCTGCTTGTGATGATGG 557
|||||
1 ThrValaspThrPheCysAlaValAlHisSerCysPheValaspGl 17
556 TAACGATGATCTGGAATATCTTAATGCTATGATGCTGCTTTGATA 507
|||||
17 yAsnGlyaspIysValGluLeuLeuAsnAlaspGlyCysAlaLeuAspL 34
506 AATATTGCTAATATATTTGAATATCCACAGATTATAGCTGGCCAA 457
|||||
34 yPheLeuLeuAsnAsnleuGluTyrrProThrAspLeuMetAlaGlyGln 50
456 GAAGCTCACGTTACAAATATCGGATGATGACAGCTTTTATCATG 407
|||||

```

```

51 GluAlaHisValTyrlYsTyrrAlaAspArgSerGlnLeuPheTyrlncY 67
406 CCAGATCAGTATTACATTAAGAAACCAATAGCGAATGTGTCGACAC 357
|||||
67 sGlnIleSerileThrIlelysgluProAsnSerIysCysGlnA1argProG 84
356 AATGTTACAGACCACAAAGGATTCGAGCTGTAAACA...GGTGGTGC 310
|||||
84 IncystrGluProGlnGlnlyPheGlyAlaIlelystrhrArgAsnGlyVal 100
309 GCAGCAAAACCTGCTGACCTGCGACACTTCGTTTACTGAAGAAAGATC 260
|||||
101 AlaAlaIaIaSerArgGlnAlaIaIaPheArgValIleuIysLysArgAs 117
259 TGCAGAACCGGAATATCATGTAGTACGACTGATATCAACACCTTG 210
|||||
117 pValArgaspGluAsnIleValaspValArgThrAspLeuAsnAlaLeuA 134
209 AATTAGCGATGATATCAAGCTTGGCAGTTGATGATTACGACCGT... 163
|||||
134 spIleAsnGluGluPheThrSerleuProAsnAlaLeuA1arghrArgSer 150
162 GCACCTTCTCAACAT...AATGACACCTGTAAATCTTGCTGACAGTACA 116
|||||
151 SerleuLeuAlaHisgluAsnGlnHisProValIleValAlaThrMetTh 167
115 AATGGAATCTGATGTCACCATTTGGCTTCGATG... 79
|||||
167 rGlnGlyIleCysMetSerileThrGlyPheThrleuA1agIyMetleuI 184
78 ..TTATGGTTTAAGCATGATGATGTCGCGCTATATT 37
|||||
184 IepheValIleValSerValAlaThrIleValAlaIleThrleu 198

```

seq_name: sp_invertebrate:096775

seq_documentation_block:

```

ID 096775 PRELIMINARY; PRT; 127 AA.
AC 096775;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)
DE CUTICULIN (FRAGMENT).
GN CUT-1.
OS Brugia malayi.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida;
OC Filarioidea; Onchocercidae; Brugia.
RN [1]
RP SEQUENCE FROM N.A.
RA LEWIS E.K., HUNTER S.J., TETLEY L., PAVIA NUNES C., BAZZICALUPO P.,
RA DEWANEY E.;
RT "cut-1-like genes are present in the filarial nematodes Brugia pahangi
RT and Brugia malayi and, as in other nematodes, code for components of
RT the cuticle.";
RL Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AJ012617; CAA10074.1; -.
FT NON_TER 1
FT NON_TER 127
SQ SEQUENCE 127 AA; 14119 MW; 3F738071 CRC32;

```

alignment_scores:

```

Quality: 662.00 Length: 127
Ratio: 5.296 Gaps: 0
Percent Similarity: 98.425 Percent Identity: 96.063

```

alignment_block:

us-09-323-427-5/rev x 096775 ..

Align seg 1/1 to: 096775 from: 1 to: 127

```

858 GTCAATTCGTTTCATTCATATTGTTTACCAAGTTGATCGTATATCG 809
|||||

```

```

1 valIleSerPheHisProLeuSerValThrLysValAlaSpArgAlaTyrAr 17
808 AGTACATCTCTTTACATGGAACCTGATAAACAGTAGTCACAGATG 759
17 gValGInCysPheTyrMetGlnAlaAlaSpLysThrValSerThGlnIleG 34
758 AGGTATCTGAATACACAACTGCTTTCAAACTCAAAATGTCCGATGCA 709
34 lValSerGlnIleThrThrAlaPheGlnThrGlnIleValPrometPro 50
708 GATATGGCGTATGAATTTTGGATGGTGACCAACCGGTCACACCATTA 659
51 ValCysArgTyrGlnIleLeuAspGlyGlyProThrGlyGlnProIleG 67
658 ATTGTGATCATTTGTCACCAAGTTATCATAAATGACATGGCATTTG 609
67 nPheAlaThrIleGlyGlnProValTyrHisLysTrpThrCysAspSerG 84
608 AAACCGTTGATCTTCTGCGCGGTGTCATTCCTGCTTGTGATGAT 559
84 lThrValAspThrPheCysAlaValAlaHisSerCysPheValAspAsp 100
558 GGAACGAGATCTGTCGGAATTTCAATGCGTATGATGATGCTGCTGA 509
101 GlysArgLysPThrValGlnIleLeuSerAlaAspGlyCysAlaLeuAs 117
508 TAAATATTGCTAAATATTTGGAAATATCCA 478
117 pLysTyrIleLeuAsnAsnLeuGlnTyrPro 127

```

seq_name: sp_invertebrate:Q21808

```

seq_documentation_block:
ID Q21808 PRELIMINARY: PRT: 290 AA.
AC Q21808:
DT 01-NOV-1996 (TREMblrel. 01. Created)
DT 01-NOV-1996 (TREMblrel. 01. Last sequence update)
DT 01-JAN-1999 (TREMblrel. 09. Last annotation update)
DE R07E3.3 PROTEIN.
GN R07E3.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
NC [1]
RP SEQUENCE FROM N.A.
RA COTTAGE A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BOWFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A.,
RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KESHAM J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PAXSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULISTON J.,
RA THERIERY-MEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RT Nature 368:32-38(1994).
RL EMBL: Z49207; CAAB9068.1;
SQ SEQUENCE 290 AA; 33215 MW; 441587FE CRC32;

```

```

alignment_scores:
Quality: 654.50 Length: 283
Ratio: 3.306 Gaps: 8
Percent Similarity: 69.965 Percent Identity: 45.230
alignment_block:

```

US-09-323-427-5/rev x Q21808 ..
Align seg 1/1 to: Q21808 from: 1 to: 290

```

891 AATCCACGCGGTATTTTGTGAACAACAACACTGTTCATTTGTTTCATGC 842
4 SerProArgGlyLeuPheLeuSerThrAsnValAlaValAlaPheAsnPr 20
841 ATATTATTGATACCAAGTTGATCGTCATATGACATGACATGCTTTTAA 792
20 oGlnPheLeuThrLysAsnAspArgValAlaPheLysValGInCysPheTyrM 37
791 TGGACCTGATAAACAGTTAGTCACAGATTGAGGTATTCGAATACAA 742
37 etGlnMetGlnArgGlnIleGlnLysValIleGlnIleSerMetPropo 53
741 ACTGCTTTCAAACTCAAAATGTCGCCGATGCCAGTATGCGGTATGAAT 692
54 ProThrMetHisSerLysGlnLeuAsnMetProValCysLysTyrGlnVal 70
691 TTGTGATGCTGACCAACCGGTCACACCTGTCATTTGCTATCATTTGTC 642
70 lLeuAspGlySerProThrGlyProProValTyrPheAlaThrValGlyG 87
641 AGCCAATTATCATTAATGACATGCGATTCTGAACCGTTGATCTTTC 592
87 lMetValTyrHisLysTrpThrCysAspThrGlnIleGlnLysThrPhe 103
591 TGGCGGTTGTCATTCGCTTGTGCGATGATGCGTACGCGGATGATCTGT 542
104 CysMetLeuValHisSerCysPheValAspAspGlyAsnGlyGlnArgVal 120
541 GGAATATTCAATATGCTGATGATGCTGCTTGTATTAATTTTGTCAATA 492
120 lGlnLeuLeuAsnAspLysGlyCysAlaLeuAspLysTyrIleLeuTha 137
491 ATTTGGAATATCCACACAGATTTTAATGCTGCGCCACAGACCTCAGTATC 442
137 snLeuGlnTyrProThrAspLeuMetAlaGlyArgGlnAlaHisValTyr 153
441 AAATATGCGGATGATGACAGCTTTTCTATCAATGCGACATGATTTTC 392
154 LysTyrAlaAspArgAspAsnMetTyrPheAspCysGlnIleSerIleTh 170
391 CATTAAAGAACCAATATGCGAA...TGTGTCGACCAACATGTTTCAGAC 345
170 rValLysGlnProGlyLeuAspTyrCysAspValProSerCysProAsp 187
344 CA..... 343
187 roProArgArgArgSerAsnThrLeuProAlaProAspAsnIle 203
342 .....CAAGATTCGAGCGTTTAAAC 320
204 ThrAlaIleAlaAlaHisIleGlnTyrGlnAspSerGlnIleIleSerA 220
319 AG...GTGGTCGCGCAGCAAAACCTGCTCCAGCTG.....CGC 285
220 sPTyrIleIleProAsnAspAlaIleIleSerLeuAsnTrpLeuGlnArg 236
284 AACTTCGTTTACTCAAGAAAGATCTGCAGAACCGAGAGATTCATTGAT 235
237 AsnPhe.....AspMetArgIleSer..... 243
234 GTACGACATGATATCACACCTTGAATTACGATGATATCAACGCTTT 185
244 ...GlnLeuCysMetThrAlaIleGlyThrThrLeuLeuValPheLeuA 259
184 GCCAGTTGATTTACGTC.....ACCGTCGACTTCTGC 153
259 snAlaPheLeuPheIleIleSerLeuValSerIleValHisValCys 274

```

seq_name: sp_invertebrate:Q21540


```

seq_documentation_block:
ID Q21540 PRELIMINARY; PRT; 550 AA.
AC Q21540:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE M142.2 PROTEIN.
GN M142.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilitida;
OC Rhabdilitina; Rhabdilitoidea; Rhabdilitidae; Peloderinae; Caenorhabdilitis.
RN [1]
RP SEQUENCE FROM N.A.
RA MCMURRAY A.;
RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KESHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERREY-MEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RT Nature 368:32-38(1994).
DR EMBL: Z73428; CA97806.1;
DR PFMW: PF00092; vwa: 1.
DR PRINTS: PR00453; VWFADOMAIN.
SO SEQUENCE 550 AA; 61445 MW; A055FB45 CRC32;

alignment_scores:
Quality: 602.50 Length: 302
Ratio: 2.953 Gaps: 7
Percent Similarity: 67.530 Percent Identity: 43.377

alignment_block:
US-09-323-427-5/rev x Q21540 ..
Align seg 1/1 to: Q21540 from: 1 to: 550

1083 GGTGAGCCAGAAATGATGATGTGACCACTTCAATCAATCAATTTTAA 1034
||| ||||| ||||| ||| : :
228 G1YthPrGluIleIleIleGysGlyProAspArgIleGlyValLysAla 244
||| :
1033 TACACGTAATGCATTCGAGACATGTTTATGGAAGGCTTTATGATC 984
||| :
244 rTlHrLysGlnProHcLndLysAsnValAlpheValMetAspHisTyrHis 261
||| :
983 AAGAGATTGCCCT.....AATGATGAGGTGAGCTCAAGT 946
: ||| ||||| : :
261 spLndLysArgAlaGlyProGluLysPheProAspSerAlaGysIle 277
||| :
945 GCCGGAATTCACCTCCATTTGATTCATGCATGTCGCGGTACACATC 896
||| :
278 ...GlyLeuHrValProPheSerAlaCysAsnValAlHisArgTyrHis 293
||| :
895 TCGAATTCACCGTGTATTTTGTAAACAACAACCTGTTGATTTCTGTTTC 846
||| :
293 rLeuAsnProLysGlyLePheValGluValSerIleValAlpheIlePheH 310
||| :
845 ATCCATTATTTGTTACCAAGTGCATGTCATATGAGTACATGCTTT 796
||| :
310 isSerLeuPheMetHrLysThrAspLndHrValLysValGlnCysPhe 326
||| :
795 TACATGGAAGCTGATAAACAGTTAGTGCACAGATTGAGTATCTGAAT 746

```

```

327 TyrMetGluAlaAspLysHisValThrValProLeuSerValSerMetI 343
|||||
745 CACAACTGCTTTCAACACCTCAATTTGTCGATGCCAGTATGCCGTTATG 696
|||||
343 eTlHrValAlpheArgLndLndLysGlnIleTyrGlnMetProGlnCysAlaTyr 360
|||||
695 AAATTTTGATGGTGGAGCCAAACCGTCACACAGTTCATTAATTTGCTATCAT 646
||| :
360 hrLeuArgLysGlyAlaProAspGlyProIleValAlrGpHeAlaThrLeu 376
||| :
645 GGTACACCACTTTATCATTAATGACATGC.....GATTCGTAAC 605
||| :
377 GlyLysrValTyrHisArgTyrGlnCysIleGluValGlnLysAlaAs 393
||| :
604 CGTTGATCTCTTCGCCGGTGTCCATTCCTGCTTTGTCGATGATGATG 555
|||||
393 pLysAspThrPheGlyMetLeuValHisSerCysTyrValAlaSpAsnGly 410
|||||
554 ACGGTATACCTGTGGAATTCCTAATGCTGATGATGATGCTGTGATTA 505
|||||
410 YrLndLysArgValAlaAspIleLeuAspSerAlaGlnCysGlyLeuAspAla 426
|||||
504 TATTTGCTAAATTAATTTGGAATATCCACAGATTTA...ATGCCGTGCCA 458
|||||
427 ValLeuLeuSerThrProAspLysArgLysThrSerLeuArgLeuAlaThr 443
|||||
457 AGAAGCTCAGCTATACAAATATGCGGATGATGATCAGCTTTTATCAAT 408
|||||
443 sProTyrHisValAlpheLysTyrLysAlaAspArgProValLeuGlnPheLnd 460
|||||
407 GCCAGATCAGTATTAACATTA...GAACCAATATGCGAATGTGTCGA 361
|||||
460 YsGlnIleThrLysCysLeuLysTyrAspGlyLysGlnGlyIleThr 476
|||||
360 CCACAATGTTTCAGAACCCAGAGATTTCGGAGCTGTAAACAGGTGATGC 311
||| :
477 Pro.....ProGlnAsnCysLysLysLeuProGluLysAspG 489
||| :
310 CGCAGCAAAACCTGCTGACCTGCGACACTTCCTTACTCAACAAAAGAT 261
||| :
489 YHlSHlSHlSHlSHlSHlSHlSHlSHlSHlSHlSHlSHlSHlSHlSHlSHl 506
||| :
260 CTGCAAGACCGGAGAAATATCATGATGTATGACAGACTGAT...ATCAACACC 214
|||||
506 euAlaAspGlyValGlyThrIleAspValAlpheThrAspSerValThrVal 522
||| :
213 CTGGA 208
|||||
523 LeuGlu 524

seq_name: sp_invertebrate:Q19053

seq_documentation_block:
ID Q19053 PRELIMINARY; PRT; 484 AA.
AC Q19053:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-JAN-1999 (Tremblrel. 09, Last annotation update)
DE E04D5.3 PROTEIN.
GN E04D5.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilitida;
OC Rhabdilitina; Rhabdilitoidea; Rhabdilitidae; Peloderinae; Caenorhabdilitis.
RN [1]
RP SEQUENCE FROM N.A.
RA MCMURRAY A.;
RL Submitted (OCT-1995) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,

```

RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RT Nature 368:32-38(1994).
 RL EMBL: 266496; CA91280.1;
 DR SEQUENCE 484 AA; 53325 MW; 6AEFC669 CRC32;

alignment_scores:
 Quality: 595.00 Length: 251
 Ratio: 3.199 Gaps: 4
 Percent Similarity: 74.104 Percent Identity: 45.020

alignment_block:
 US-09-323-427-5/rev x Q19053 ..

Align seg 1/1 to: Q19053 from: 1 to: 484

```

1083 GGTGAGCCAGAAATTTGATGACCACTTACATACATCAATTTTAA 1034
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
14 G|G|u|p|p|r|o|l|u|v|a|l|C|y|S|g|u|t|h|r|a|l|S|e|r|i|e|u|l|e|u|r|h|e|l|y 30
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1033 TACACGTAATGCATTCGAGACATGTTTATGTGAAGTCTTTATGATC 984
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
30 s|h|r|a|g|s|e|r|p|h|e|a|n|s|l|y|S|v|a|l|p|h|e|v|a|l|l|y|S|g|l|T|Y|v|a|l|S|e|r|g 47
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
983 AAGAAGTTCGCGTAATGATGACGAGTGCATGACGTCGCGCAATTTCA 934
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
47 l|u|p|r|o|S|e|r|C|y|S|m|e|t|h|r|v|a|l|G|l|y|S|p|l|y|S|h|r|g|l|y|h|S|a|r|p|h|e|g|l|u 63
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
933 CTTCCATTTATTCATGCATGATGTTGGCGGTACACATCTCGATCCACG 884
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
64 V|a|l|a|r|g|l|S|a|S|p|e|r|C|y|S|g|l|v|a|l|a|r|g|l|a|r|g|l|u|l|e|a|S|n|... 78
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
883 TGGTATTTTGTAAACAACAGCTGTGTCATTTGTTTCATTCATTTATG 834
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
79 .g|l|v|a|l|v|a|l|l|e|S|e|r|a|l|h|r|v|a|l|l|e|l|e|S|e|r|p|h|e|l|S|e|r|l|e|h|e|l 95
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
833 TTTCCAAAGTTCGTCGATGCATGACGATGATGACGTCGTCGTCGTCG 784
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
95 l|e|h|r|l|y|S|l|e|S|p|h|a|r|g|a|l|a|r|y|a|r|v|a|l|S|e|r|C|y|S|p|h|e|r|y|v|a|l|G|l|u|g|l|y 111
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
783 GATAAACACTTAGTCACATGATGAGTATCTGAATCAACAAGTCTT 734
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
112 T|h|r|l|y|S|v|a|l|h|S|a|S|n|h|S|v|a|l|S|p|l|e|S|e|r|a|l|e|u|r|h|r|g|l|u|l|e 128
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
733 TCAAACTCAATTTGCCGATGCCAGTATGCCGTTATGAATTTTGATG 684
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
128 u|l|e|u|r|g|l|S|e|r|G|l|u|h|r|o|l|e|u|r|v|a|l|C|y|S|a|r|g|l|y|G|l|u|l|e|l|e|a|S|n|g 145
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
683 GTGACCAACCGGTCAACAGTTCATTTGCTATGATGTCGTCGTCGTCG 634
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
145 l|u|a|a|..g|l|y|l|S|e|r|p|r|o|l|e|u|r|l|e|u|r|a|r|g|l|l|e|g|l|y|a|S|p|r|g|l|u|v|a|l 160
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
633 TATCATTAATGACATCGATTCGTAACCGTGTGATCACTTTCGCGCGT 584
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
161 T|y|h|S|l|y|S|r|h|r|C|y|S|v|a|l|a|G|l|u|l|e|G|l|a|S|n|V|a|l|y|r|C|y|S|m|e|t|y 177
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
583 TGTCCATTCCTGCTTGTGATGATGCT..AAGCGTACTGCTGGAAA 537
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
177 S|v|a|l|h|S|e|r|C|y|S|h|r|v|a|l|y|r|a|S|p|l|y|G|l|u|r|p|r|o|v|a|l|h|r|v 194
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
536 TTTGTAATGCTGATGATGCTGCTGATTAATATTTGCTAAATTAATTG 487
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
194 a|l|l|e|S|p|a|l|a|S|n|g|l|y|S|e|r|v|a|l|S|p|l|y|v|a|l|l|e|u|r|G|l|a|S|n|l|e|u 210
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

486 GAATATCCACAGATTTTAATGCTGGCCAGACGTCACGATATACAAATA 437
      ||||| :|||:||||| |||||: ||| :|||:|||||
211 G|l|u|t|t|h|r|e|r|S|p|e|r|h|u|r|a|l|a|G|l|y|S|v|a|l|p|r|o|v|a|l|h|e|u|r|h|e|l|y|S|p 227
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
436 TCGGATCGATCAGACAGTTTTCATCATCAATCCGATGATGATTTACATTA 387
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
227 e|a|l|a|S|p|l|y|S|a|l|a|G|l|y|e|r|h|p|h|e|a|S|n|C|y|S|g|l|l|e|G|l|n|l|e|u|r|h|l|e|l 244
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
386 AAGACCAATATACGAAATCT.....GTTGACCAAGATGTTGACGAACA 343
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
244 Y|a|S|p|a|l|a|S|n|T|Y|G|l|y|C|y|S|e|r|a|S|h|r|h|r|o|l|e|G|l|n|C|y|S|p|r|o|h|r|S|e|r 260
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
342 CAA 340
      |||
261 G|l|n 261

```

seq_name: sp.invertebrate:Q93532

seq_documentation_block:
 ID Q93532 PRELIMINARY; PRT; 315 AA.

```

AC Q93532:
DT 01-FEB-1997 (TReMBLrel. 02, Created)
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
DT 01-JAN-1999 (TReMBLrel. 09, Last annotation update)
DE F20D1.8 PROTEIN.
GN F20D1.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;
OC Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA BURTON J.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RP
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,
RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  

  elegans.";  

  RT Nature 368:32-38(1994).  

  RL EMBL: 278542; CAB01742.1;  

  DR SEQUENCE 315 AA; 35785 MW; CD2BAB3D CRC32;

```

alignment_scores:
 Quality: 578.50 Length: 281
 Ratio: 3.029 Gaps: 5
 Percent Similarity: 67.972 Percent Identity: 40.214

alignment_block:
 US-09-323-427-5/rev x Q93532 ..

Align seg 1/1 to: Q93532 from: 1 to: 315

```

1152 CGTCTATATGCTTTCGTGACTACACTTATTCGATGCTCTTATGATTC 1103
      ||||| |||||
30 A|g|l|l|e|r|p|r|o|l|e|u|r|h|e|u|r|C|y|S|y|S|.....T|y|h|S|a|l|a|l|u|G|l 41
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1102 GGTGTAACATGCTGTGGAAGTGAGCCAGAAATGATGATGACCAACTT 1053
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
41 n|l|e|a|S|p|a|S|n|g|l|y|G|l|u|r|p|r|o|l|e|u|r|l|e|a|r|g|S|g|l|y|S|e|r|g|l|u|S 58
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1052 C|A|T|T|A|C|A|T|C|A|T|T|T|A|T|A|T|A|C|A|G|T|A|T|G|C|A|T|T|G|A|G|A|C|A|T|G|T|T|A|T 1003
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```



```

176 eLeuYstrYrAlaLysValGlyAspGlnValAlaHisArgTrpGlnCysA 193
      ::::::::::||| ::||::: ||| ::::::::::|||
614 ATTCTGAACCGGTGATCTTCTGGCGGCTGTCCATCCGCTTCTTC 565
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|
193 spSerGlu.....AspTrgGlyLeuLeuValHisSerCysTrpVal 206
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
564 GATGATGTAACGCGTGAATCTGTGAATTTCTAAATGTCGATGATGTC 515
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
207 GluAspGlyGlnGlyLysGlnMetIleIleAspGlnArgGlyGlysh 223
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
514 TCTTGATTAATATTGCTTAATTAATTTGGAATATCCACAGATTTA..A 468
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
223 sThrAspArgLeuLeuGlyAspProThrTrpValGluAlaLeuAsnM 240
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
467 TGGCGGCCAGAGAGCTACGATACAAATATGCGGATGATCAGCGTT 418
      ||||| ::|::|::|::|::|::|::|::|::|::|::|::|::|
240 eValAlaTrpArgGlnSerPheValPheLysPheAlaAspTrgIleAla 256
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
417 TCTATCAATGCCAGATCAGTATTACCATTAAGAACCAATAGCGAATG 368
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
257 ArgPheGlnCysGlnLeuArgLeuGlyLeuLysAsp..AspGlyGlyCy 272
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
367 T.....GTTGCACCAATGTTTCAGAACCCAGAGATTGCGAGCTG 327
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
272 sAspGlyIleThrProPrometCysSer.....P 282
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
326 TTAAGAACAGGTGGTCCGCGACCAACCTGCTGCAGCTGCGCACTTGT 277
      ||| ::|::|::|::|::|::|::|::|::|::|::|::|::|
282 heLysAspRlaAsnSerAsnGlnLysGlnIleValLysArgAsnAlaTrg 298
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
276 TTAATCAAGAAAAGATCTGCAGAACCGGAGAAATATCATGATGATGAC 227
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
299 ThrPheLysProLysAspAlaAsp.....MetPheSerGlnThrVal 312
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
226 TGATATCAACACCCCTTGAATTAAGCATGATTAAGCTTTGCCGCTTG 177
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
312 LysTrpAlaMetAspRlyGlnAsnGlnLysAspSerThrSerAlaGlnAla 329
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
176 ATTTCAGTCACCGTGCATTTCTGCAACATTAATGACACCTGTAAATCT 127
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
329 LulLeuArgLul..... 332
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
126 GCTGCAGTACAAATGGAATCTGCATGTCACCA.....TTTGG 89
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
333 ..LeuAspProGlnThrIleCysLeuAlaProLysLeuLeuValAla 348
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
88 CTTCGTAATGTTTATGCGTTTAAGCATTGATGATGCGCGCATTA 39
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
348 aValAlaThrPhePheThrPheIlePheValLeuPheValIleThrIleLeu 365
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
38 TTACCATTTTCGTTTAAATTTCTGTCACAAATCAG 7
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
365 aValValAlaTrpHisArgTrpCysLysAsnGln 375
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
seq_name: sp_invertebrate:O9XVN2
seq_documentation_block:
ID_O9XVN2 PRELIMINARY; PRT; 610 AA.
AC_O9XVN2:
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DE F53B6.6 PROTEIN.
GN F53B6.6
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita;
OC Rhabdilita; Rhabditoidea; Rhabdilitidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WHITE S.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

```

```

RT elegans.";
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,
RA CRAWTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULLSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans ";
RL Nature 368:32-38(1994).
DR EMBL; Z81086; CAB03116.1; -.
SQ SEQUENCE 610 AA; 68585 MW; D793181D CRC32;

alignment_scores:
Quality: 511.00 Length: 270
Ratio: 2.762 Gaps: 6
Percent Similarity: 68.519 Percent Identity: 37.407

alignment_block:
US-09-323-427-5/rev x O9XVN2 ..
Align seg 1/1 to: O9XVN2 from: 1 to: 610

1134 ACTACATATTGATGATGCTATTATTCG.....ATTCGGTTGACAAATG 1091
||| ||||| ||||| ||| ::|::|::|::|
61 ThrIleuIleLeuLeuSerThrSerCysPheGlnIleGlnAsnI 77
1090 TGTGCAAGTGAACCCAGAAATTTGATGTGGACCACTTAATTAACATGA 1041
||| ||||| ||||| ::|::|::|::|
77 yValAlaGlyLysProGlnValAlaPheCysGlyIleAspRliLeuArgVal 94
1040 ATTTCATACAGTAATGATGATGAGAGACATGTTATGTAAGAGTCTT 991
::|::|::|::|::|::|::|::|::|::|::|::|::|
94 yValAlaSnThrGlnHisProPheAsnGlnLysArgIleTrpValAspGly 110
990 TATGATCAAGAAGTTCG.....CGTAATGTAAGTGAAGTCAAGT 947
|||::|::|::|::|::|::|::|::|::|::|::|::|
111 SerAspLysGlnHisCysValGlnHisSerAlaAspAlaHisSerSer 127
946 TGCCGGAATTCACATTCGATTCATGATGCAATGTTGGCGGTACAGAT 897
::|::|::|::|::|::|::|::|::|::|::|::|::|
127 oGlnGluPheThrIleProIleGlyAlaCysAsnMetAlaArgGlnAlaTrg 144
896 CTTGTAATCAGCTGATTTTGTGTAACAACAACCTGTTCATTCGTTT 847
::|::|::|::|::|::|::|::|::|::|::|::|::|
144 TrIleuHisProArgGlyIleSerPheSerPheThrMetIleHisSerPhe 160
846 CATTCATTAATTTGTTTACAAAGTTCATGTCATATGAGTACAAATGCTT 797
|||::|::|::|::|::|::|::|::|::|::|::|::|
161 HisProPhePheValThrGlyMetAlaAspArgAlaPheSerIleArgGlysh 177
796 TTACATGGAAGCTGATAAACAGTTAGTGCACAGATGAGATGATCGAAA 747
|||::|::|::|::|::|::|::|::|::|::|::|::|
177 ePheLeuGlnSerIleLysGlyLeuAsnAlaGlnIleAspValGlyThrL 194
746 TCACAGCTGCTTTTCAACCTCAAAATGTCGCCGATGCGCATGCGCTTAT 697
::|::|::|::|::|::|::|::|::|::|::|::|::|
194 euAlaProGlnHisValAspGlnGlyTrpSerLeuProValIleCysAlaTrg 210
696 GAATTTTGGATGCTGGAACCAACCGTCAACAGTTCAATTTGCTATCAT 647
::|::|::|::|::|::|::|::|::|::|::|::|::|
211 HisLeuLysAsp...GlyTleGlnGlnHisValLeuArgPheAlaGlnVa 226
646 TGCTCAGCCAGTTTATCATTAATGACATGCGATTTCTGAACCGTTGATA 597

```

[illegible]

```

seq_documentation_block:
ID      093967      PRELIMINARY;      PRT;      440 AA.
AC      093967; 094405;
DT      01-FEB-1997 (TREMBLrel. 02, Created)
DT      01-NOV-1998 (TREMBLrel. 08, last sequence update)
DT      01-JAN-1999 (TREMBLrel. 09, last annotation update)
DE      ZK265.8 PROTEIN.
GN      ZK265.8.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
RN      Rhabditina; Rhabditoidea; Rhabditidae; Peloiderinae; Caenorhabditis.
RP      [1]
RP      SEQUENCE FROM N.A.
RA      DOBSON R.;
RL      Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE; 94150718.
RA      WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA      BONFIELD J., BUTTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA      CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA      GARDNER A., GREEN P., HAKINS T., HILLIER L., JER M., JOHNSON L.,
RA      JONES M., KESHAW J., KISTEN J., LAISTER N., LATRILLE P.,
RA      LIGHTING J., LOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA      PARSONS J., PERCY C., RIKEN L., ROOPA A., SAUNDERS D., SHOWNKEEN R.,
RA      SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SUSTON J.,
RA      THERRY-MEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA      WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLMAN P.;
RT      "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT      elegans.";
RL      Nature 368:32-38(1994) .
RN      [3]
RP      SEQUENCE FROM N.A.
RA      LENNARD N.;
RL      Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR      EMBL; Z81143; CAB03520.1; -.
DR      EMBL; Z75713; CAB03520.1; JOINED.
DR      EMBL; Z75713; CAB00057.1; -.
DR      EMBL; Z81143; CAB00057.1; JOINED.
SQ      SEQUENCE 440 AA; 49299 MW; 3F2BBA2 CRC32;

```

```
alignment_scores:
  quality: 477.50
  length: 320
```

```

Ratio: 2.296          Gaps: 9
Percent Similarity: 65.000      Percent Identity: 33.438

alignment_block:
US-09-323-427-5/rev x Q93967  ..

Align seg 1/1   to: Q93967 from: 1   to: 440

```

1155 ATTGCGATTATGCTTTGCTGCTACCTACATATGCGATTGCTGCTATTGCGAT 1100
4 LcUvSLlELeuIlleIllelAtrPhrIcglUyTrarGvAlaAlaAlaIl 20
1105 TCCGCGTTCGCAATGCGTGTGCAAGTGCAGCAGCAAAATTAAATGCGACCA 1056
20 eSerlIleAspAnslUlellelGlyGlnProAspIleGlyLucSyLeuGlnA 37
1055 GTTCGAATACAACTCAATTTTAAATACACGTAATGCATTGCGAGCAGATTT 1006
37 spGlnIleArgIleTrValLysThrArgLysIlePheAlaGlyArgIle 53
1005 TATGTGAAGAAGCTTTATGATCAAGAGGTGCCCGTAAATGAGAGGTGG 956
54 TGTAlaLysGlyArgAlaGlnLeuGlnsPcSYTLvSAspAspPheIl 70
955 AGCTACAGTATGCC.....GGAATTTCACCTTCATTTGATTGCATTCACA 915
70 yAnslGlnLysThrArgLysProHlsPheAspLeuGlnPheGlyAlaCysG 87
914 ATGTTGCGCGTACACAGATCTGTGAATCCACGTGATTTTGTGAACACA 865
87 lYmetLysSerIleuArgSerValAspProArgGlyMetetylGlyIle 103
864 ACAGTTGCTATTGCTGTTATCATCATTTATTTGTACCAAAAGTACGCG 815
104 ThrValValValSerPheHlsProLeuPheIleThrLysValAspGlnAl 120
814 ATATCAGATACATGCTTTTACATGCGAACCTGATAAACAGTTAGTGCAC 765
120 aTyHlsValLysCysPhePheGlnGlnAlaHlsLysGlyLeuThrAlaG 137
764 AGATTAGGATATGTGAATACACACTGCTTTTCAAACTCAAAATTGTCCG 715
137 lUeudlYalSerMetIleProThrThrGlnLeuGlnAlaArgHlsGly 153
714 ATGCCAGTATGCCGATGAAT.....TTGGA 686
154 lIleProGlyCysThrLysSerIleHlsArgSerThrIleAspGlnLeu 170
685 TGGTGA...CSAACGCGTCAACACAGTTCAATTTGCTATCATTTGCTGAC 639
170 palagIyArgProAlaGlyAsnValIleGlnPheHlsArgValGlyGlnA 187
638 CAGTTATCATAAATGGACATGCGATTCTGAAACCGTTGATCTTTCTGC 589
187 rGValLeuHlsGlnIleThrHlsCysAsnAspGln.....MetYrGly 200
588 GCGGTGTGCATTCGCGCTTTGCGATGTGAATGTAACGCGATACGTGA 539
217 pAlaIleAspAspLysGlyCysProIleAspProIleLeuIleThrGlyI 234
488 TGGAAATATCCACAGATTTAAATG...GCTGGCGAAGAACGTACGATATAC 442
234 lArgIlyIleSerIleAspLeuGlnArgAlaTyAlaGlnSerIleValPhe 250
441 AAATATGCGGATGCGATACAGGCTTTTGCATTCATCAACGACAGATAGATTAC 392
251 LysPheAlaAspLysProGlyAlaIleThrPheIleCysGlnValGlnMetCys 267
391 CATTTAAAGAACCAATAGCGAATGT.....GTTCGACCCACAAATGT. 352

```

267 smetLysLys...HisGlyMetCysAspGlyIleThrProProSerCysG 283
      :::::  :::::  |||
      351 .....TGAGAACCAAGAGATTGGAGCTGTAAACAGGTGGCC 310
            ||:::  :::::
      283 LysMetSerArgValIleSerValGlyIleuAspAsnGlyGlyPhe 299
            :::::
      309 GCAGCAAAACCTGCTGCAGCACTCGTTTACTCAAGAAAGATC 260
            :::::
      300 GluGluGluIleuLysAlaProSerSerArgArg.....LysThrTh 313
            :::::
      259 TGCAGAACCG 250
            :::::
      313 rProLysPro 316

seq_name: sp_invertebrate:Q23097

seq_documentation_block:
ID Q23097 PRELIMINARY; PRT; 387 AA.
AC Q23097;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JAN-1999 (TReMBLrel. 09, Last annotation update)
DE W01A8.3 PROTEIN.
GN W01A8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditidae; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP MEDLINE: 94150718.
RA WILSON R., AINSWORTH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRANFIELD M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SUDSTON J.,
RA THERRER-MEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RT Nature 368:32-38(1994).
DR EMBL: Z71267; CAA95850.1; -
SQ SEQUENCE 387 AA; 43212 MW; D127FF23 CRC32;

alignment_scores:
Quality: 476.00 Length: 360
Ratio: 2.235 Gaps: 8
Percent Similarity: 59.167 Percent Identity: 31.389

alignment_block:
US-09-323-427-5/rev x Q23097 ..
Align seg 1/1 to: Q23097 from: 1 to: 387

1077 CCAGAAATGATGATGGACCACTCAATCAATCAATTTTATATACAGC 1028
      ||:::  :::::
48 ProLysValIleCysAlaGluAsnAspLeuAlaLeuAspIleValThrSe 64
      :::::
1027 TAATGATTCGAGAGACATGTTATGTGAAGGCTTTATGATCAAGAG 978
      :::::
64 rLysProPheArgIleAsnIlePheValIysGlyArgAlaLysAspLys 81
      :::::
977 GTTGGCGTATGATGAGAGTGCATCAATGCCGGAATTCACATTCGA 928
      :::::
81 ertCysArgInserTyrAlaAsnAsnGlyThrAsnSerTyrSerLeuPro 97

```

```

927 TTTGATTCATGCAATGTGGCGGTACACGATCTCTGAATCCAGCTGTAT 878
      :::::
98 LeuGlyLysCysGlyMetGlnArgLeuArgSerAlaAsnProAlaGly 114
      :::::
877 TTTTGAACAACAACATGTGTGCTATTCGTTTCATCCATTA...TTTGTGA 831
      :::::
114 LAsnPheMetValThrValIleValSerPheHisProAlaGlyPheIle 131
      :::::
830 CCAAAAGTTGATCGCATATCGAGTACATGCTTTTACATGGAAGTGAT 781
      :::::
131 hLysAsnAspArgAlaPheHisValLysCysPheTyrMetGluProAsp 147
      :::::
780 AAAACAGTTAGTCACAGATTGAGTATGATGATGATGATGATGATGAT 731
      :::::
148 GluIleValThrGlnAsnIleAspValSerMetIleProThrIleGlu 164
      :::::
730 AACCTAAATTTGCCGATGCCGATATGCCGTTATGAATTTTGATGATG 681
      :::::
164 userAspSerMetValMetProLysCysGluTyrSerValArgArgAsp 181
      :::::
680 GACCAACCGGTACACAGTTCAATTTGCTATCATCTTGTCAGCCAGTTTAT 631
      :::::
181 LProAsnGlyProThrLeuThrTyrAlaAsnValGlyAspIleValPhe 197
      :::::
630 CATAAATGGACATCGCATTCGAAACCGTTGATACCTTCTGCCCGGTGT 581
      :::::
198 HisValThrProLysCys.....ThrProAlaAspMetGlyMetLeuVal 211
      :::::
580 CCATTCCTGCTTTGTGATGATGATGATGATGATGATGATGATGATGAT 531
      :::::
211 LysLysCysPheValThrAspLysPheValGlyIleAsnHisAlaValAla 228
      :::::
530 ATGCTGATGATGATGCTGCTTGAATATTTGCTAATATTTGATGATAT 481
      :::::
228 sPheAspGlyCysAlaThrAspProPheLeuSerGlyLeuSerTyr 244
      :::::
480 CCAAGCAATTTAATG...GCTGCCAAGAAGTACAGCTATACAAATATGCA 434
      :::::
245 AspAlaSerLeuMetArgAlaHisAlaSerSerGlnValPheLysTyrAl 261
      :::::
433 GGATCATCACAGCTTTTCTATCATATGCCAGATCATATATACATTAAG 384
      :::::
261 aspSerAsnGlnLeuTyrPheThrCysGlnIleAlaGlyLeuCysGlnLys 278
      :::::
383 AACCAAT.....ACGGAATGTTGCGACCAAGATG..... 352
      :::::
278 lMetGlyMetCysGlnGluValThrProProAsnGlyValLys 294
      :::::
351 .....TCAGAACCAAGAGATTGGAGCTGTAAACAGGTGGTCCGC 308
      :::::
295 LeuLeuSerGlnLysAlaSerGlyAspGly..... 303
      :::::
307 AGCAAAACCTGCTGCAGCTGCCAAGCTGTTTACTCAAGAAAGATCTG 258
      :::::
304 ..... 307
      :::::
257 CAGAACCGAGATATATGATGATGATGATGATGATGATGATGATGATGAT 208
      :::::
307 ysrArgIleValAspArgSerAspTyrGluIleAspValAlaThrSerGlu 323
      :::::
207 ATTAGGATGATTAATCAAGCTTTGCCAGTTGATTTAGCTACCGTGCATC 158
      :::::
324 Leu.....LeuValLeuAspProAlaAspArgGlyLe 334
      :::::
157 TCTGCACATATATGACAACTGTAACTACTGCTGACAGTACAAATGGA 108
      :::::
334 uLeu..... 335
      :::::
107 TCTGCATGTACCATTTGGCTTCTCAATGTTTATGGGTTTAAAGCATTCGA 58
      :::::
336 ..AlaProSerProPheCysValProArgLeuLeuLeuProValLeuPro 351

```

57 TTGATTGCTGCCGCTCATTTATACCATTCG 28
 |||||
 352 LeuileuileuilerleuileuileuSerLeuThr 361

seq_name: sp_invertebrate:Q22680

seq_documentation_block:

ID Q22680 PRELIMINARY; PRT; 609 AA.
 AC Q22680;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE T22C8.7 PROTEIN.
 GN T22C8.7
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA THOMAS K.;
 RL Submitted (Apr-1995) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A.,
 RA CRAWFORD M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KESHAV J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTEIN L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 DR EMBL; 249071; CAA88879.1; .
 DR HSSP; P03000; 1TIF.
 SQ SEQUENCE 609 AA; 68882 MW; 0BA2EE0C CRC32;

alignment_scores:

Quality: 453.50 Length: 287
 Ratio: 2.399 Gaps: 11
 Percent Similarity: 65.854 Percent Identity: 37.282

alignment_block:

US-09-323-427-5/rev x Q22680 ..

Align seg 1/1 to: Q22680 from: 1 to: 609

1159 GATGATTCGCTTATGCTTCTGCTACTACCTTATTCATTCGCTTATT 1110
 |||:|||||:||||| ||| |||||:|||||
 20 AspAsnSerAsnTyrAsnPro...TyrTyrThrHis.....Me 31
 1109 CGATTCCCGTTGACAAATGGTGTGCAAGGTGCGCAAAATTTGAATGGA 1060
 |||:|||||
 31 tglu.Pro.....ProlysLeuGluCysGly 39
 1059 CCAACTTCATAACAATCAATTTTAATACAGTAATGCATTCGAGAGACA 1010
 :|||:|||||:|||||:|||||:|||||
 40 SerGluGlyLeuArgLeuHisIleAsnProThrGlyThrPheGlyGly 56
 1009 TGTATTATGCAAGAAGTCTTTATGATCAGAAGGTGCCGTAATGATGAAG 960
 |||:|||||:|||||:|||||:|||||
 56 sValTyrValArgGlyPhePheProGlnThrValCysHisLeuAsnTyrC 73
 959 GTGAGAGT.....CAAGTGGCCGAATTCACATTCATTT...GAT 922
 |||:|||||:|||||:|||||:|||||
 73 ystThrArgLeuThrAsnArgProIleValMetAspLeuProPheArgGly 89
 921 TCATGCAATGTTGCGGCTACACGATCTTGAAATCCAGTGTATTTTGT 872

||||| ||| |||:||||| ||| :|||
 90 ProCysAsnValArgArgArgAsnValAlaProProSerIleSerTy 106
 871 AACAAACACTGTTGCTCATTTGCTTTCATCATTCATTTATTTGTTACCAAGTTG 822
 :|||:|||||:|||||:|||||:|||||
 106 rAspValThrValIleIleGlnHisIleProLeuPheValThrSerPhe 123
 821 ATGTCATATCGAGTACAAATGCTTTTACATGCAAGCTGATTAACAGTT 772
 |||:|||||:|||||:|||||:|||||
 123 spLysAlaTyrArgLeuAsnGlyIleTyrArgGlnGlnGlnSerThrLeu 139
 771 AGTCACAGATGAGGTATCTGAATC...ACAAGTCTTTTCAAACTGA 725
 :|||:|||||:|||||:|||||:|||||
 140 GlnGlnArgIleAsnValSerAspIleProSerThrAlaLeuGlnSerIy 156
 724 AATTGCCCGCATGCCGATGCGCTATGCAAAATTTGGATGCTGGACCAA 675
 :|||:|||||:|||||:|||||:|||||
 156 sAsnAlaProlys.....CysArgTyrAspValLeuSerGlySerLeu 171
 674 CCGGTACACCAAGTTCAATTTGCTATCATGCTGCGGCGGTTTTCATTA 625
 :|||:|||||:|||||:|||||:|||||
 171 snGlyProValValArgPheAlaAsnValGlyAspValValValHisIys 187
 624 TGGACATGCGATTCGAAACCGTTGATACTTTCTGCGGCTGTCATTC 575
 |||:|||||:|||||:|||||:|||||
 188 TrpThrCysAspSer.....AspArgPheGlyPheValValHisSe 201
 574 CTCCTTTTCGATGATGCTACCGGTGATACTCTGCAAAATTTCTAAATCGTG 525
 |||:|||||:|||||:|||||:|||||
 201 rCysValValArgAspGluSerGlyLysAspPheGlnPheIleAspGlu 218
 524 ATGATGCTGCTTGTGATTAATATTGCTAAATATTTGGAATATCCACA 475
 |||:|||||:|||||:|||||:|||||
 218 rGlyGlyValThrAspPheSerLeuPheProGluValSerTyrSerAsp 234
 474 GATTTAATGCTGCGCAAGAAGCT...CACGTATACAAATATGCGGATCG 428
 |||:|||||:|||||:|||||:|||||
 235 AspLeuLysSerAlaPheThrAlaValArgAlaPheArgTyrAlaAspG 251
 427 ATCACAGCTTTTCTATCAATGCCAGATCAGTATTTACCATTAAGAACA 378
 :|||:|||||:|||||:|||||:|||||
 251 nValMetValHisPheSerCysGlnIleThrThrCysGlnLysGlnGlu 268
 377 ATAGC...GAATGTGTTCCGACCAATGTTTCAGAACCCACAGAGATTGCGA 331
 |||:|||||:|||||:|||||:|||||
 268 snGlyCysGlnGlyIleSerProIleCysArgProMetAspLeuGly 284
 330 GGTGTTAA 322
 :|||:|||||
 285 ProlLeuLys 287

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 16, 2000, 04:55:03 ; Search time 4841.33 Seconds
(without alignments)
-728.156 Million cell updates/sec

Title: US-09-323-427-5

Sequence: 1 tgcctctgatttgacgaa.....gcataagaagacatcatcat 1161

Scoring table: IDENTITY_NDC
Gapop 10.0 , Gapext 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_bal:*
2: gb_da2:*
3: gb_on:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl2:*
10: gb_pl2:*
11: gb_pl2:*
12: gb_ro:*
13: gb_ro:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_vl:*
18: em_fun:*
19: em_hum1:*
20: em_hum2:*
21: em_in:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vl:*
32: gb_hlg1:*
33: gb_hlg2:*
34: gb_hlg2:*
35: gb_in1:*
36: gb_in2:*
37: em_bal:*
38: em_da2:*
39: em_hum3:*
40: gb_pl4:*
41: gb_pl4:*
42: gb_hlg3:*
43: gb_hlg5:*
44: gb_hlg6:*

45: gb_hlg7:*
46: em_hlg1:*
47: em_hlg2:*
48: em_hlg3:*
49: em_hum5:*
50: gb_pl3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	339.2	29.2	25823	CEF22B5	250044 Caenorhabdi
2	323.8	27.9	34831	CEC47G2	249125 Caenorhabdi
3	315	27.1	2310	CELCUT1	M55997 C.elegans c
4	261.4	22.5	724	AF125580	AF125580 Nucleare
5	249	21.4	5792	ALU073005	U73005 Ascaris lum
6	217.2	18.7	2584	MAATCUT1	X96677 M.artifellia
7	190.6	16.4	637	BMA012617	AJ012617 Brugia ma
8	156.6	13.5	358	BPA012618	AJ012618 Brugia pa
9	107.4	9.3	39478	CEB3FE1	281088 Caenorhabdi
10	83.8	7.2	31536	CEE04D5	266496 Caenorhabdi
11	74.8	6.4	32412	CEP53B6	281086 Caenorhabdi
12	67	5.8	39736	CEB07E3	249207 Caenorhabdi
13	64.2	5.5	110000	CEY39A1_3	Continuation (4 of
14	62.6	5.4	36306	CEM142	273428 Caenorhabdi
15	62.2	5.4	36069	CEZK265	281143 Caenorhabdi
16	56.2	4.8	25284	CEM01A8	271267 Caenorhabdi
17	54.8	4.7	28396	CELF10E7	U41264 Caenorhabdi
18	53.6	4.6	40265	CEM28	249911 Caenorhabdi
19	52.8	4.5	34368	CEM06D12	282073 Caenorhabdi
20	51.8	4.5	40600	CEP20D1	278542 Caenorhabdi
21	51.8	4.5	164548	CEY70D2	ALU08880 Caenorhab
22	49	4.2	23953	CELR06A1	U23449 Caenorhabdi
23	49	4.2	31731	CELR03H9	U21318 Caenorhabdi
24	47	4.0	45713	CELB0511	AF067608 Caenorhab
25	46.8	4.0	27770	CEP22C8	249071 Caenorhabdi
26	43	3.7	164399	PFMA13P6	298551 Plasmidum
27	42.6	3.7	165302	AC009451	AC009451 Homo sapi
28	41.2	3.5	37872	CEI04F8	265655 Caenorhabdi
29	41.2	3.5	102195	CEY71H9	AL021575 Caenorhab
30	41.2	3.5	110000	CEY102E5_0	AL022276 Caenorhab
31	40.2	3.5	82652	ATAC007020	AC007020 Arabidops
32	40.2	3.5	95713	AF085279	AF085279 Arabidops
33	40	3.4	9901	2 U67530	U67530 Methanococc
34	39.8	3.4	12976	CEY53H1B	AL132851 Caenorhab
35	39.8	3.4	110000	CEY53H1_2	Continuation (3 of
36	39.4	3.4	5214	AF090533	AF090533 Dictyoste
37	39.4	3.4	6100	AB017910	AB017910 Dictyoste
38	39.2	3.4	91894	AC005739	AC005739 Homo sapi
39	39.2	3.4	124181	AF207067	AF207067 Homo sapi
40	38.8	3.3	5038	SGU40026	U40026 Streptococc
41	38.8	3.3	35955	CEP53B7	272510 Caenorhabdi
42	38.2	3.3	2296	AF044022	AF044022 Eurytides
43	38	3.3	190000	AC004479	AC004479 Homo sapi
44	37.8	3.3	1310	EGU21005	U21005 Euglena gen
45	37.8	3.3	11163	2 AE001500	AE001500 Helicobac

ALIGNMENTS

RESULT 1
CEP22B5
LOCUS CEF22B5 25823 bp DNA
DEFINITION Caenorhabditis elegans cosmid F22B5, complete sequence.
ACCESSION Z50044
VERSION Z50044.1 GI:899234
KEYWORDS HTG: Cuticulin; Elongation factor; GTP-binding ADP-ribosylation

CDS

gene

gene

CDS

Query Match	Similarity	29.2%	Score 339.2	DB 34	Length 25823
Best Local	Similarity	63.3%	Freq. No. 41e-70		
Matches	633	Conservative	0	Mismatches 233	Indels 134
				Gaps	2
OY	169	acgtaatacgaactgagcaagcttgatatacaccgcgaattcaagaagtggtgatatacag	228		
Db	3064	ATGAACACCATTGTTGACACACTTACATTTTCCTCCAAACTCCAAAGTTGTGAGCTCAGC	3123		
OY	229	tcgtatcaatcagatattcttcgcggtctgcgagatctttcttgatataaagaattgcg	288		
Db	3124	ACGGACATCAAGAAATATTTCTCATTACGAGACGGATCTCTTTCCAAAACTCTGAGAGA	3183		
OY	289	agctcagcagagtttctgtcgcgcacacacgttcttaacagctccgaatcccttgcgtc	348		
Db	3184	CTGACACAGTTTGTTAGCTTGC-----TTAACTGCTCCAAATCCTGTGGGCTC	3231		
OY	349	tgaatatgtgtgcgaacaacattgcatttggttctttaaagtgaaatacgaatctgcga	408		
Db	3232	AGAGCAGCTGGGGCAGGACACTCTCTCATTTGGTTCTTAAAGTAATGGAATCTGGCA	3291		
OY	409	ttgatagaaaacgttgatcgatccgcacatttgtatacgttaagctctctggcagccat	468		
Db	3292	TTGATAGAAAAGTTGAGAACGATCAGGCTATTTTGAGACGTGAGGCTCTGTGCCAGCAT	3351		

QY	469	taaacctcgttggataltccaaattatttagcaaatattatcaagaagcatccatcagc	528
Db	3352	CAAAATCAGTGGATATTCCAAATTTGTTGAGCAAGAACTTATCAAGGACCAACCGTTCTC	3411
QY	529	atttgaatttccacagatataccggtttaccatcattcgcagaagccggaaatgtaacaacgc	588
Db	3412	ATCCAAAATTTTGCAGGTACACCATTTTCATCATCAACAGTGCAAAGATGAACGCACAC	3471
QY	589	gcagaagaatcatcaacggttcagaatcgcagtgttccattatgatataaacgycgtgcacat	648
Db	3472	GCAGAAAGATCAACAGGTTTCAGAAATCAAGGTCCATTGTTGTATAGACTGTTGTTCAT	3531
QY	649	gataagcaaatgaactcgtgttgaacgggttgcgcaccatccaaatttcaaacgycatac	708
Db	3532	AGTTTCGGAATTAACATGGTTCTTCACGTTTGCTCCTCGTTAAGAAATCCTATTTACAGAT	3591
QY	709	tggcatcgggacaatttgaatttgaagaagcagttgattatagatacctcaactctgtgc	768
Db	3592	TGGCATTTGGAAACACCTGGGTTTGGAACACGGTGTGAAGATCAGAGACCTTCATTTGGGT	3651
QY	769	actaaccttattatgaactcgtatcattgataaagaatgtactgatatgc-----	818
Db	3652	ACGACATGCTTGTGTGGCTTCATGTAGAGAGCAATTCGAACACCGTAGGCACCTGAAATTCA	3711
QY	818	-----	818
Db	3712	ATTTCGATTTTTCAAAGCAAGCTTCTAGTTGATATCAAGCTTTATAATTTCGGATTC	3771
QY	818	-----cgatcaact	826
Db	3772	TAAACCTTCAAAGACATTCCTTAAAAATTACAGAAATTTCATAGCTTACCGATCAAC	3831
QY	827	ttgttgtaacaataatagatagaagaacgaatagcaacacagttgtgttacaataataaccagt	886
Db	3832	TTGGTGCACAAATTTGGGGATTTGGAAGAAGACAGACAGATGTTGTGTAAACGAAGACTCCCTTT	3891
QY	887	ggattccagaagatcgttgaacgcacaatcgtatgaataccaatgggaagttaaatccgcga	946
Db	3892	GGATTCCAGGATCTGTGTACGACCAACGTTGCAGATTCGAATGGAAAGTTCGATTCACAGG	3951
QY	947	acttgaagtcacacttcatcattacgcgaacacttctgtatcataaagaccttccatatca	1006
Db	3952	ACTTGAACGACCTCTTCGTCATTTCTACATTCCTTGTGTATCAAAAAGTCCCTTAACGTAG	4011
QY	1007	acatgctccttgaatagtcattacgtgtatataaattgatcg 1046	
Db	4012	ACGTGTCCCTCGAAGGCATTTAGCTGTGTGAAGTTTACG 4051	

RESULT 2	CEC47G2/c				
LOCUS	CEC47G2	34831 bp	DNA	INV	02-SEP-1999
DEFINITION	Caenorhabditis elegans cosmid C47G2, complete sequence.				
ACCESSION	Z49125				
VERSION	Z49125.1	GI:790365			
KEYWORDS	HTG; Cuticulin; Fork head.				
SOURCE	Caenorhabditis elegans.				
ORGANISM	Caenorhabditis elegans.				
	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilitida; Rhabdilitina; Rhabdilitidae; Rhabdilitidae; Peloderinae; Caenorhabditis				
REFERENCE	1 (bases 1 to 34831)				
AUTHORS	Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Cosey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Larrelle, P., Lightning, J., Lloyd, C., McMurtry, A., Mortimore, B., O Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopa, A., Saunders, D., Shownkeen, R., Smaildon, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J. and Wohlman, P.				

TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans

JOURNAL Nature 368 (6466), 32-38 (1994)

MEDLINE 94150718

REFERENCE 2 (bases 1 to 34831)

AUTHORS Palmer, S.

JOURNAL Direct Submission

Submitted (27-APR-1995) Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rwenematode.wustl.edu

COMMENT Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.

For a graphical representation of this sequence and its analysis see: -

http://webpace.sanger.ac.uk/cgi-bin/displaydb-wormacc/class-sequence/object-C4762

Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

IMPORTANT: This sequence is not the entire insert of clone C4762. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone C4762 is at 1 in this sequence. The true right end of clone C4762 is at 6388 in sequence 249912.

The true left end of clone T24F1 is at 34731 in this sequence. The true right end of clone T05B9 is at 12750 in this sequence. The start of this sequence (1..101) overlaps with the end of sequence 249129.

The end of this sequence (34731..34831) overlaps with the start of sequence 249912.

Location/Qualifiers

1..34831

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="III"

/clone="C4762"

2562..4878

/gene="cut-1"

join(2562..2679,2992..3217,3898..4636,4687..4878)

/gene="cut-1"

/note="similar to cutcullin: cDNA EST EMBL:C10444 comes from this gene; cDNA EST EMBL:C12078 comes from this gene; cDNA EST yk262c12.5 comes from this gene; cDNA EST yk262c12.5 comes from this gene; cDNA EST yk265h3.5 comes from this gene; cDNA EST yk278g8.3 comes from this gene; cDNA EST yk439e8.3 comes from this gene; cDNA EST yk439e8.5 comes from this gene; cDNA EST yk380a11.3 comes from this gene; cDNA EST yk380a11.5 comes from this gene; cDNA EST yk380a11.5 comes from this gene"

/codon_start=1

/protein_id="CA88934.1"

/db_xref="GI:3875030"

/db_xref="SWISS-PROT:Q03755"

/translation="MTWKPIICLALVLSASAIIVDNNVEGEPEVECPNSIYVNT RNPEGHVYVGLYDOAGCRSDEGRQVAGLEPFDSCNTAFRTSLNKGVEFTVV ISFHQVETKVDRAVRIQCIFYMESDKTYSTQIEVSDLTATFQTVPMVPVKEYLLDG GPGGPIQFATIGOOVYHKMTCDSETTDFCAVYHSGVDGNGDTVOILNEGCALD KELLNNEFYPTDIAAGGEAAHYKKAADRSQLYCOISTITTCRPGSECAPIKSPGCG GAVQAGAGAAHAAAPQAGVEVQAAVAHAAAPVAHAAAPVAPRATTLQRL LRKRSFGENGILDRVREINTLIDIMEGASAPAEAAALVSESVRRATSTGICLTP IGFASFGLGIGTIVTALSATIFYVAPRTSHK"

15433..17108

/gene="C4762.2"

join(15433..15770,16205..16531,16772..17108)

/gene="C4762.2"

/note="similar to fork head domain protein; cDNA EST EMBL:T01618 comes from this gene; cDNA EST EMBL:D70725 comes from this gene; cDNA EST EMBL:D66805 comes from this gene; cDNA EST yk268f8.3 comes from this gene; cDNA EST yk268f8.5 comes from this gene; cDNA EST yk474g11.5 comes from this gene; cDNA EST yk637f8.3 comes from this gene; cDNA EST yk670a12.3 comes from this gene"

/codon_start=1

/protein_id="CA88935.1"

/db_xref="GI:3875031"

/db_xref="SPTRMBL:Q18694"

/translation="MLFSMESLSTSTPKLEPPPKLEPEVTINQVYDLPBSNRLSE PSTASVLEHLDKPGESKRKRSRIGDEPTEDDEPVPRKAKRNHSTSSAADSSDA KDDDDDDSTSRKSMGSKRSKSHAKPYSTALIAMSLNSPEKRLTSEICETINK FEYKKEKPDYONSIIRNLNLDCEVAVARPGPGNGWALDPNCEMDNCSFLR RRRYKRNSTPDYHEHMSHPMPRPPLPGGPRPPRPMPANPMGHNMPAVN MPATFQUNDSOKILSMASRIIMPAPVSSGCKRTSSSSPENNCSAVSKLSA" complement(24417..26914)

/gene="C4762.4"

complement(join(24417..24975,25043..25176,25237..25522,25668..25907,25967..26054,26106..26227,26277..26608,26681..26785,26846..26914))

/gene="C4762.4"

/note="similarity to Trichostrongylus colubriformis 11 kd secretory protein (Swiss Prot accession number P21937); cDNA EST EMBL:D3349 comes from this gene; cDNA EST EMBL:D37644 comes from this gene; cDNA EST EMBL:D36149 comes from this gene; cDNA EST EMBL:C11456 comes from this gene; cDNA EST EMBL:C13631 comes from this gene; cDNA EST yk358b10.3 comes from this gene; cDNA EST yk358b10.5 comes from this gene; cDNA EST yk301f4.3 comes from this gene; cDNA EST yk295f2.3 comes from this gene; cDNA EST yk295f2.5 comes from this gene; cDNA EST yk550d1.3 comes from this gene"

/codon_start=1

/protein_id="CA88936.1"

/db_xref="GI:3875032"

/db_xref="SPTRMBL:Q18695"

/translation="MGTISLAVQLFIVFLTSLYLNKSTIRKONPIVITSTIGWY SLIVFVLDPDAITFPFKCENDROILINTTSPVAVPECEKPGVDDVLFNLM RVVYVADLTWILPLLOSIVTAGNFTLIGKIAAYINNAIYATISLGFALIIA MEKVSAINIENLKIYVYASNTWGLFLLVLLGALVEPLRSIMHNRHRYLRKTF DIEKLASEKSEAEENVKDIYKKVYVENSKNDSNGCRKRYRTLSKFSDDVDNLFPS ROVIDNAHLDSESGSEAKLISLHKRTIYAVQTLNATQWKLVDRLALENLAY ESNGYNLELSRNTCPVIGVRFVWYRTQIPEVTEFVEFVSECFEVSYS TISPAAVTEYASRPFHYKQTOFAFGIYVILITSAYFTFRLQIYVYHLDPGHDY ENSLFSAIILCRITPILCNFLGIMHDSIISAKSGFIEOTKILMGHLDPILIA KGINIYIPICIIICAIHYRYGVAYVYHNGFQDFQFVADENINMISGRSIVQIEN SIKRSNRSQRTQWNTNSFGSSNAGNSTTSKFRSKNKNDEERMLDDEDEVEEST RISMSPTREHSSSGFFDDM" 30076..34049

/gene="C4762.5"

join(30076..30134,30184..30492,31050..31274,31321..31480,31534..31653,31701..31835,31889..32104,32160..32570,32639..32935,32998..33215,33263..33651,33802..34049)

/gene="C4762.5"

/note="cDNA EST CEMS75F comes from this gene; cDNA EST EMBL:C13621 comes from this gene; cDNA EST yk227d8.5 comes from this gene; cDNA EST yk257e7.5 comes from this gene; cDNA EST yk319c4.5 comes from this gene; cDNA EST yk331d4.5 comes from this gene; cDNA EST yk335b12.5 comes from this gene; cDNA EST yk486d7.5 comes from this gene"

/codon_start=1

/protein_id="CA88937.1"

/db_xref="GI:3875033"

/db_xref="SPTRMBL:Q18696"

/translation="MAAEGRKRNKRKRSRDRREGTSASSDAYSSDAEDIDNODDD PMLNLDSLDHNVSEARPEPDLDIVHSPPNMEPDNODLEHBEKAFBEGSGSR IAOESASVADHNGRSDSDSHTRARSLCYHSPNSAPAKLSKFTSIFQMFVAKKEE ENSLMLKTDNFTLEDVLLNFEVQESRQKAEVQYITISRENMKALLESLSLPKIN TDLPMKOYRSLFIASLFTIRGTVDVOKQVYVTEFRCKIVDLNKPCTLNHLYVAF FAKIMECLSHRDPULTFSLOETKFPFKCLRNINLGAIECLLENLVAIRPTSEBTRIV


```

Db 1300 TCACAGCTGGCATGTGTAGACCTGGCTGTGAGAGCAGTGGAGATCGGAGACTTCGA 1241
Qy 762 tctgtgaactactgtttatcacgttcacatgtaaaagatg 804
Db 1240 TCTGTGTGACACAGCTTGTGTGGACTCCATGTAGAGCATGG 1198

RESULT 4
AF125580 724 bp. mRNA INV 05-MAY-1999
LOCUS Muchereria bancrofti cuticulin-1 mRNA, partial cds.
DEFINITION AF125580
ACCESSION AF125580
VERSION AF125580.1 GI:4741874
KEYWORDS
SOURCE Muchereria bancrofti.
ORGANISM Muchereria bancrofti.
Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida;
Filarioidea; Onchocercidae; Muchereria.
REFERENCE 1 (bases 1 to 724)
AUTHORS Ramzy, R., Helmy, H., Adely, M., Curtis, K. and Weil, G.
TITLE Muchereria bancrofti L3 cuticulin-1 cDNA partial sequence
JOURNAL Unpublished
AUTHORS Ramzy, R., Helmy, H., Adely, M., Curtis, K. and Weil, G.
REFERENCE 2 (bases 1 to 724)
TITLE Direct Submission
JOURNAL Submitted (03-FEB-1999) Research & Training Center on Vectors of
Diseases, Ain Shams University, Abassia Square, Cairo, Egypt
FEATURES
source
location/Qualifiers
1..724
/organism="Muchereria bancrofti"
/db_xref="taxon:6293"
/dev_stage="L3 larvae"
/country="Egypt; Nile Delta"
<1..623
/note="similar to Ascaris lumbricoides cuticulin-1"
/codon_start=3
/product="cuticulin-1"
/protein_id="AAD28743.1"
/db_xref="GI:4741875"
/translation="VDFPCAIVHSCFVDNGSKVELLNADGALDKFLLNNLEYPT
DLMAQGEAHVKKYADRSLFYQCOISTIKRPNKSCORPCTEGRGATITRNQVAA
ASRQAAARVLRKRDENVDTVDITDINALDINDEFSLNLRHSLSLAHENGHP
VIVATMTGIGISITGFTLACMLIFVIVATVATVATVATLLRSHSTRV"
BASE COUNT 248 a 112 c 128 g 236 t
ORIGIN

Query Match 22.5%; Score 261.4; DB 35; Length 724;
Best Local Similarity 68.2%; Pred. No. 8.7e-52;
Matches 396; Conservative 0; Mismatches 176; Indels 9; Gaps 2;

Qy 37 aataatgcgcgcgaacatcatgcttaaacccataaacatgtgagaaccaaagtg 96
Db 581 AATTGTAGACAGAAACATATACAAATATTAGCATACCAGCAAGTAAACCGGTAT 522
Qy 97 tgacatgcagatccatcttctgactgcgcgaagtaagctgtccatctatgtg--- 154
Db 521 TGACATACATATACCTTGTGTGATGTGTGACACATTTACCGGATACCATTTCTCATGTGC 462
Qy 154 ---cgaagtgcaagtgtaagtaaacactgcgcgaagctgataatcatcgtcaattc 210
Db 461 TAACATATGATGATCGATGCGTAGCGCATTTTGTAATGAGTAACCTCTTCAATTATATC 402
Qy 211 aaggtgtgtatataatgctgacataatgatatctcccgcttctgcgaatcttctc 270
Db 401 CAATGCATTAAAGATGGTACGTACATCAACAATATTTTATCTTACATCTCTCTTTT 342
Qy 271 gagttaaacgaagtgcaagctgcgcgaagctgttctgcgcgacaa-----ccgtgtttaa 327
Db 341 CAATCTCTTAATGACAGACGCTTGTGCGATGACACGTGACACACCATTTGGCGGTTTAAT 282
Qy 328 agctcgaatccctgtgtgtctgaaacatgtgtgtcgaacacatcgcattatgtgtc 387

```

```

Db 281 TCGGCCAAACACCTTGGGTGGCTTACGTGATGTGTGTGACATTTACTATTTGGTTCCT 222
Qy 388 aatgtaaacacgcatctggtcattgataaagaagctgtgacatccgcacatcttctaaac 447
Db 221 AATTGTATGCTAATTTTGCATTTGATTAATAAAGTTGTGTGATCATGCGCTATTGTATAC 162
Qy 448 gtcagcttcttgcgcgaagccatnaatctgttgcataatccaaatattatagcaaatatt 507
Db 161 GTGTGCTTCTTGCCAGCATTAATCTGTGATATCTTAATATTATTAAGTAGAAATTT 102
Qy 508 atcaagagacatccatcagcatcttagaatctccacagatataccgltacatcatcgac 567
Db 101 GTCCAAAGCAGCAACCATGACATTTAATGACTTACTTTATCCATTTACATTCATCAAC 42
Qy 568 aagcaggaatgacacaccgcgcgaagaatcatcaacggtt 608
Db 41 AAAGCATGATGTACACCGCACAGATGTATCAACTGTTT 1

RESULT 5
ALU73005 5792 bp DNA INV 14-AUG-1997
LOCUS ALU73005/c
DEFINITION Ascaris lumbricoides CUT-1-like cuticulin protein precursor
ACCESSION (ascut-1) gene, complete cds.
VERSION U73005
KEYWORDS U73005.1 GI:1657624
SOURCE
ORGANISM common roundworm.
Ascaris lumbricoides
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilia; Ascaridida;
Ascaridoidea; Ascarididae; Ascaris.
REFERENCE 1 (bases 1 to 5792)
AUTHORS Timouni, M. and Bazicalupo, P.
TITLE cut-1-like genes of Ascaris lumbricoides
JOURNAL Gene 193 (1), 81-87 (1997)
MEDLINE 97390131
REFERENCE 2 (bases 1 to 5792)
AUTHORS Timouni, M. and Bazicalupo, P.
TITLE Direct Submission
JOURNAL Submitted (01-OCT-1996) GA3, I.I.G.B., via G. Marconi, 10, Napoli,
NA 80125, Italy
FEATURES
source
location/Qualifiers
1..5792
/organism="Ascaris lumbricoides"
/variety="sum"
/db_xref="taxon:6252"
766..881
/gene="ascut-1"
/join(766..954,2851..3239,3583..4083,4499..5081)
/gene="ascut-1"
/product="CUT-1-like cuticulin protein AscUT-1"
766..5081
/gene="ascut-1"
766
/gene="ascut-1"
/note="site of transplicing by the SL1"
882..929
/gene="ascut-1"
/join(882..954,2851..3239,3583..4083,4499..4693)
/gene="ascut-1"
/note="cuticlin gene; ASCUT-1; homologous to cuticlin
proteins of other nematodes"
/codon_start=1
/product="CUT-1-like cuticulin protein precursor"
/protein_id="AAB66646.1"
/db_xref="GI:1657625"
/translation="MCRVAVSLALFLGLAAIPVDNGVEGEPELEGPTSTIVNFRN
PFEHAYVKGLYDDEGCRSDGGRVAGISLPFSCNVARTRSLNPGIFVTTTVIS
FHLPLITKVRVAVVOCFEMADGKDTSTOIEVSTTAPOTQIYPMPCVYEILDGP
TGQPIGTATTCQGYTHKWTCDSEIVDFPCAVVHSCFVDDSGDITQLINEGCLMDY
LNLLEYPTDLMAGQEAHVKKYADRSLFYQCOISTIKRPNKSCORPCTSEPGCFGA
VRPGSIAPKRRCQRLRIKSSGDVNDLITDPSALDISDRDEALPMDLHRAR

```

Query Match	21.4%	Score 249;	DB 34;	Length 5792;
Best Local Similarity	69.7%;	Pred. No. 7.7e-49;		
Matches 352;	Conservative 0;	Mismatches 150;	Indels 3;	Gaps 1;
mat_peptide	HARCOOYLSPANGICMSPGCFSTFWCLAAVLAALAAVYVVSFKLRPOKA" . join(930). .954,2651. .3239,3583. .4083,4499. .4690)			
3' UTR	/gene="ascut-1" /product="cut-1-like cuticlin protein" 4694. .5081			
polyA_signal	/gene="ascut-1" 5064. .5070 /gene="ascut-1"			
BASE COUNT	1689 a	1254 c	1213 g	1636 t
ORIGIN				
Query Match	21.4%	Score 249;	DB 34;	Length 5792;
Best Local Similarity	69.7%;	Pred. No. 7.7e-49;		
Matches 352;	Conservative 0;	Mismatches 150;	Indels 3;	Gaps 1;
OY 199	atcgtaattccaagaggggtgatctcagttcgtatcatcaatgaatattccggttcgc	258		
DB 4080	ATCACTAATGTCGAGGGCGCTGAATCCGGTGTATCAATAGGGGTGTCTCAATGACACC	4021		
OY 259	agatcttctcttgatgaacgaagatctgcagcgtcgaagcaggtttgtctgc--ggacc	315		
DB 4020	CCCACTCTTTTGTGATCAGAGCGAGGTTGGCAGCGCGCTCTTTTCGGCGCATGAGGCG	3961		
OY 316	acctgttttaacagctccgaatccttctgtgtctcgaacatgtgtgtcgaacatcgtc	375		
DB 3960	AACCTGGAGCGACAGCCCAAAATCCTTGGGCGCTCACTGCAAGTGGGTGGGCAATTCGCT	3901		
OY 376	attgtgtctttaatcgtgaatcagtcagtcagtcagtcagtcagtcagtcagtcagtc	435		
DB 3900	GTTTGGCTCTTTTATGTGTATGTATGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGT	3841		
OY 436	atattgtatacgttagcttctgtgcagcacaataatcgtgtgtatattccaaatatt	495		
DB 3840	ATATTTGTAGACGTGCGCTTCTTGGCGGACATCAAAATGCGGGGTACTCGAGGTTGTT	3781		
OY 486	tageaataattatataagaagacatcatcagcatttagaatttccacagatcacggt	555		
DB 3780	CAGCGGATTTTGTCAACGCGCAAAACCCCTCCCTGTTGAGGATCTGGATGTATCAACCGCT	3721		
OY 556	accatcatcgaacaaagcggaatggaacacgcggaagaagatataaggtttcgaatc	615		
DB 3720	GCATCATCAAGAAAGCATGAGTGAACGACACACAGACGATGCACAGTTTCGAAATC	3661		
OY 616	gcatttccattatgataaacatcgtctgacccaatgataagcaaatgaactcgtttcagct	675		
DB 3660	GCAAGTCACACTGTGTGTATCTTGTCTGACCCAGTGTGTGTGTGTGTGTGTGTGTGTGT	3601		
OY 676	tgtgtccaccatccaattcataa 700			
DB 3600	TGGACCTCCGTCGAGAAATCTGTAAA 3576			
RESULT 6				
MAMTCUT1/c	2584 bp	DNA	INV	25-NOV-1997
LOCUS				
DEFINITION	M.artiellia Mcut-1 gene.			
ACCESSION	X96677			
VERSION	X96677.1	GI:2648040		
KEYWORDS	cuticle protein; cuticlin 1; Mcut-1 gene.			
SOURCE	Melioidogyne artiellia.			
ORGANISM	Eukaryota; Metazoa; Nematoda; Secernentea; Diplogasteria; Tylenchida; Tylenchina; Tylenchoidea; Heteroderidae; Melioidogyninae; Melioidogyne.			
REFERENCE	1 (bases 1 to 2584)			
AUTHORS	De Giorgi,C.			
TITLE	Direct Submision			
JOURNAL	Submitted (15-MAR-1996) C. De Giorgi, Dipartimento di Biochimica e Biologia Molecolare, Via Orabona 4, 70126, Bari, ITALY			
REFERENCE	2 (bases 1 to 2584)			
AUTHORS	De Giorgi,C., De Luca,F., Di Vito,M. and Lambertini,F.			
TITLE	Modulation of expression at the level of splicing of cut-1 RNA in			

JOURNAL	MOL. Gen.	Genet.	253 (5),	589-598 (1997)
MEDLINE	97218031			
FEATURES	Location/Qualifiers			
SOURCE	1..2584			
	/organism="Meloidogyne artiellia"			
	/db_xref="taxon:42426"			
	/clone_1lb="lambda7-2(1)"			
	/clone="pUC (3000) "			
	/clone="mp19(1500) "			
	join(<467..943,1139..1735,1825..>2052)			
mRNA	/gene="Mtcut-1"			
	467..2052			
gene	/gene="Mtcut-1"			
	<467..943			
exon	/gene="Mtcut-1"			
	/number=1			
CDS	join(467..943,1139..1735,1825..2052)			
	/gene="Mtcut-1"			
	/codon_start=1			
	/product="cuticlin-1"			
	/protein_id="CA65452.1"			
	/db_xref="GI:2648041"			
	/db_xref="SPRMBL:O18479"			
	/translation="MRKLFGVFLNALMFYVAIPVDNGVEGEDELICGPTSTIV NFNRNPFEGHVVVKGLFDQAGCRSDERGVAVIELPDCNVARTAEKGFVS TVLSFHPDFTKVDRAYRVCGFMEDKRTSALESEITOTOTOJOPVPVCKEIE LEGALGPLOAFITAQVOYTSRGCDSETIDTPAVVSHCVDDNGNDTVOILNEGSC ALDFLLNNLEYFPDLTAGDGRANHYKKADRSLTYCCQISTTIKEPSHCARPCAEE SGFAVNAVGGAGAGAPPPAAAPATPPPAPAPAPAPAPAPVORIRIQLRLRR RRAARGQVLDVTRTEFNALDISDKFPNLPAIRRHHLMEILSAEPORRFYYGGGOOV CMSEFRANHLRYCHCNGCSHMWVSHVSTCLFEVFESNEIK"			
intron	944..1138			
	/gene="Mtcut-1"			
	/number=1			
exon	1139..1735			
	/gene="Mtcut-1"			
	/number=2			
intron	1736..1824			
	/gene="Mtcut-1"			
	/number=2			
exon	1825..>2052			
	/gene="Mtcut-1"			
	/number=3			
POLYA_signal	2518..2523			
BASE COUNT	761 a 639 c 559 g 625 t			
ORIGIN				
Query Match	18.7%; score 217.2; DB 34; Length 2584;			
Best Local Similarity	56.8%; Pred. No. 2.4e-41;			
Matches	581; Conservative 0; Mismatches 243; Indels 198; Gaps 2;			
Db	285 gcgcagcctgcacgcaagtcttgcctgcggcacccatcgttttaaacagctccgaattcgtg 344			
Db	1547 GCGGTGGGGACACGCCGCACACGCTCCACCGGGCCGACATTGACGGCGGTGAAGCCGCTTG 1488			
Oy	345 gtctcgacaacttgctgcgaacacacttgcctatttgcttctaatagttaactgatct 404			
Db	1487 GCTCAGCACACTTTTGAGCAGACCATCACTGTGGTCCTTGATTGATTATACGATCT 1428			
Oy	405 ggcattgataagaagcctgtgataccgaatacttatctgatacgtgagctcttgccag 464			
Db	1427 GGCAATTGGTAGACAGCTTGAGAGCATCCGCCATTATTGTAGACGTGAGCCTTTGGCCGG 1368			
Oy	465 ccattaaactcgttgatatalaccaaatatattagcaaatattatcatcaagaagcacatcat 524			
Db	1367 CAGTCAAAATCGGTTGGATTCCAAAGTTGTTAGACAAGACTTGTCCAGGCACAGCTTC 1308			
Oy	525 cagcatttagaattccacaglatcacccgttacatccatctcgacaagaaggaaatgagcaa 584			
Db	1307 GTTCTTCAAAATTTGACAGCGGTGTCCCATTCCTTCATCCACACAGAGCTGTGCACAA 1248			

Oy	585	ccggccgagaaatatacaagcgtttcagaaatcgcatccttaattgatgaacacgctgcac	644
Db	1247	ccggccgagaaagtgtcaaatcgtctcagactcacaggctccacttggtaactgttgccg	1188
Oy	645	caatgatagaacttgaacttgcttgacgggttggctccacatccaaat-----	694
Db	1187	caatggtgscgaattgaatgggttgcccaagtcggcgcccttcggagatctttggaagaa	1128
Oy	694	-----	694
Db	1127	taaaattcaattaaaaattattttgtttgttttttactaaataaaagtttaaaaaacaaca	1068
Oy	694	-----	694
Db	1067	tttttgagcatctttaaaccttgsgcttagacgaacaaaattgttaaaatcaactcgatt	1008
Oy	694	-----	694
Db	1007	ttatgcgaattcaaaagtttgcatatgtatttactgaagaagtagagctcttctttcttac	948
Oy	694	-----ttcataacggcctactgcatcgggacaaatttgagttgaaaaagcagttgtaatt	749
Db	947	ttacctcaattttggcacacaggcatagggacacaccttgaggcttggaactgttgtagatct	888
Oy	750	cagatatacctcaatctgctgactcaactggtttatcagcttcctcattgataaagcattgac	809
Db	887	cgagacacctcgagctgtgcccgaacggcttctgtctcttcattatgaagacctggacac	828
Oy	810	gatatgcacatcaactctgtgtaacaabaatlgatlgaaacgaatagacacagttgttg	869
Db	827	ggttagggcccggtcaacttgggtgacaaactgtgatggaagagatgaccaccgctgtgg	768
Oy	870	ttacaaaaataccacgttgatcagaagatcgtgtaacgcaaatgtcatgaatcaaatg	929
Db	767	agacgaacactctcttggggtt---cagcatcaatggcggaacattgacaaagctcgaacg	711
Oy	930	gaagtgaaattccggcaacttgcacacttcaacatcatgaagcaaccttgcatact	989
Db	710	gaagctcaattcccgcaacttgcggcgctgctcattccgaacggcattccggccttgctca	651
Oy	990	aaagaccttcacataaacaatgctccttgaatgacgtglatataaattgatatgta	1049
Db	650	agagccctttccacatagacgtcccccgcamaaggggttgctgctgaattgacagtgca	591
Oy	1050	ttgaagttgctcacatctaattctcgtgcaccttcgacacatgtgaacccgaatcg	1109
Db	590	ttgagcttgggccacatttaattctctggctcgcttcacacaccgttgctcactggaattg	531
Oy	1110	aa 1111	
Db	530	ca 529	
RESULT 7			
BMA012617/c			
LOCUS	BMA012617	637 bp	DNA
DEFINITION	Brugia malayi cut-1 gene, partial.		
ACCESSION	AJ012617		INV
VERSION	AJ012617.1	GI:3859854	29-JUN-1999
KEYWORDS	cut-1 gene; cuticlin.		
SOURCE	Brugia malayi.		
ORGANISM	Brugia malayi.		
	Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;		
	Filarioidea; Onchocercidae; Brugia.		
REFERENCE	1 (bases 1 to 637)		
AUTHORS	Lewis,E., Hunter,S.J., Tetley,L., Nunes,C.P., Bazzicalupo,P. and		
	Devaney,E.		
TITLE	cut-1-like genes are present in the filarial nematodes, Brugia		
	pahangi and Brugia malayi, and, as in other nematodes, code for		
	components of the cuticle		
JOURNAL	Mol. Biochem. Parasitol. 101 (1-2), 173-183 (1999)		
MDLINE	993939397		

REFERENCE	2 (bases 1 to 637)
AUTHORS	Devaney,E.
TITLE	Direct Submission
JOURNAL	Submitted (06-NOV-1998) Devaney E., Veterinary Parasitology, University of Glasgow, Bearsden Road, Glasgow, G61 10H, UK
FEATURES	Location/Qualifiers
source	1. .637
gene	/organism="Brugia malayi"
exon	/db_xref="taxon:6279"
exon	/clone="cut-1"
exon	1. .484
exon	/gene="cut-1"
exon	1. .164
exon	/gene="cut-1"
exon	/number=1
exon	join(<2. .164,266. .>484)
exon	/gene="cut-1"
exon	join(<2. .164,266. .>484)
exon	/gene="cut-1"
exon	/codon_start=1
exon	/product="cuticlin"
exon	/protein_id="CAI10074.1"
exon	/db_xref="GI:3858955"
exon	/db_xref="SPTREMBL:O96775"
exon	/translation="VISEFHLSTYKRVBRAYRVOCFYEADKTVSTQIENVSEITTAQOTQIVMPVQREIILDDGPTGPIQFATIGOPVYHKMTCDSIVDFPFCVAVHSCFEVDGN
exon	GDVIELISADSCALDKYLLNNLEYP"
exon	165. .265
exon	/gene="cut-1"
exon	/number=1
exon	266. .484
exon	/gene="cut-1"
exon	/number=2
BASE COUNT	174 a 136 c 134 g 193 t
ORIGIN	
Query Match	16.4%, Score 190.6; DB 34; Length 637;
Best Local Similarity	68.3%; Pred. No. 4.5e-35;
Matches 334; Conservative	0; Mismatches 54; Indels 101; Gaps 1;
472 atctgttggaatcccaaatatattagaaatattatccaagagacatcgaacatt 531	
489 ACCGTGCGGATATTCCTCAATATTATTCAGCAAGATATTATTCGAGACACACCATCTGCACCT 430	
532 tagaatlccacaglatcacgcgtltaaccatcatcgacaaagcaggaaatgagacaacgcgcgcga 591	
429 CAGAAATTTCCACCGCATACCGTTGCCATCAWCCACAAAGCAGAGTGGACACATGCGCA 370	
532 gaaatlatcaacgcgtttagaatcgcatgctcatcttataagataaactgctgaaccaatgat 651	
369 GAAGGTATCAACGGTTTCGGAATCAACAGGTTCATTTGGATTAACCTGGTTGGCCAAATGGT 310	
652 agcaaatgaaactggtltacacgcgtltgacccacatcccaaat----- 694	
309 AGCAATTTGAATTTGGCTGCTCCGGTTGGTCCACCATCCAAATCTAAATATTGTGCACCA 250	
694 ----- 694	
249 TCTACTATGTGATAAATAGATTTTTCATGTTTGATTTGAATTTGATAAATTAATGCAAAAT 190	
694 -----ttcataaagcgatattgcatcggaacgaatttgat 730	
189 TCACAAAATTCACAAAAGCATGCTCTGTATCGACAAACAGGCGATGGGACAAATTTGAGTT 130	
731 tgaaaagcagtlgtlgtatllcagataaccacaactcgtgtgacacaaactglttllcagactlcc 790	
129 TGAANAAGCAGTTGTGATTTTCAGACACCTCAATCTGGGGTGTGACCGCTTTTTCAGCTTCC 70	
731 atgtcaaacgcatlgtactcgcgatatgacgacatcaactlgtgttaacaataatgagatgaac 850	
69 ATATGAAGACACTGCCTCGGTATGCTCGATGCAACTTTTGTGACGAGACAGTGTGGAAG 10	


```

Qy      851 gaaatgaca 859
        |||||
Db      9 GAAATTACA 1

RESULT 8
LOCUS   BPA012618      358 bp      DNA
DEFINITION Brugia pahangi cut-1 gene, partial.
ACCESSION AJ012618
VERSION   AJ012618.1 GI:3858956
KEYWORDS cut-1 gene; cuticlin.
SOURCE   Brugia pahangi.
ORGANISM Brugia pahangi
          Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida;
          Filarioidea; Onchocercidae; Brugia.
REFERENCE 1 (bases 1 to 358)
AUTHORS   Lewis, E., Hunter, S.-J., Tetley, L., Nunes, C. P., Bazzicalupo, P. and
          Devaney, E.
          cut-1-like genes are present in the filarial nematodes, Brugia
          pahangi and Brugia malayi, and, as in other nematodes, code for
          components of the cuticle
          Mol. Biochem. Parasitol. 101 (1-2), 173-183 (1999)
JOURNAL 9939397
MEDLINE 2 (bases 1 to 358)
REFERENCE Devaney, E.
AUTHORS   Direct Submission
TITLE     Submitted (06-NOV-1998) Devaney E., Veterinary Parasitology,
          University of Glasgow, Bearsden Road, Glasgow, G61 1QH, UK
FEATURES
          source
          1..358
          /organism="Brugia pahangi"
          /db_xref="taxon:6280"
          join(<L..191,350..>356)
          /gene="cut-1"
          1..191
          /gene="cut-1"
          /number=1
          1..356
          /gene="cut-1"
          join(<L..191,350..>356)
          /gene="cut-1"
          /note="unnamed protein product"
          /codon_start=1
          /protein_id="CAI10075.1"
          /db_xref="GI:3970656"
          /db_xref="SPTREML:096776"
          /translation="PIQFATIGQPVYHKWTCDSFTVDFCAVHSCFVDDGNGDVEI
          LNDGCLDRLYLLNNLEYPYDL"
          192..349
          /gene="cut-1"
          /number=1
          350..356
          /gene="cut-1"
          /number=2

BASE COUNT 110 a 64 c 64 g 120 t
ORIGIN
Query Match 13.5%; Score 156.6; DB 34; Length 358;
Best Local Similarity 87.7%; Pred. No. 4.7e-27;
Matches 171; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```

```

Db      75 AAAGGTATCAACGGTTTCGGAATACAGGTCATTTGTGATTAACGTGGTGGCAATGGT 16
Qy      652 agcaattgaaactg 666
        |||||
Db      15 AGCAATTGAATTGG 1

RESULT 9
LOCUS   CEF53F1      39478 bp      DNA
DEFINITION Caenorhabditis elegans cosmid F53F1, complete sequence.
ACCESSION Z81088
VERSION   Z81088.1 GI:1627965
KEYWORDS HTG.
SOURCE   Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
          Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
          Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 39478)
AUTHORS   Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,
          Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
          Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
          Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
          Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
          Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
          O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, A.,
          Saunders, D., Showkeen, R., Smaildon, N., Smith, A., Sonhammer, E.,
          Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
          Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
          Wilkinson-Spratt, J. and Wohldman, P.
          2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
          elegans
          Nature 368 (6466), 32-38 (1994)
JOURNAL 94150718
MEDLINE 2 (bases 1 to 39478)
REFERENCE Burton, J.
AUTHORS   Direct Submission
TITLE     Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail:
          joesanger.ac.uk or rwenematode.wustl.edu
          Coding sequences below are predicted from computer analysis, using
          predictions from GeneFinder (P. Green, U. Washington), and other
          available information.
          For a graphical representation of this sequence and its analysis
          see:
          http://webace.sanger.ac.uk/cgi-bin/display?db=wormace&class=Sequence&object=F53F1
          Current sequence finishing criteria for the C. elegans genome
          sequencing consortium are that all bases are either sequenced
          unambiguously on both strands, or on a single strand with both a
          dye primer and dye terminator reaction, from distinct subclones.
          Exceptions are indicated by an explicit note.
          IMPORTANT: This sequence is NOT necessarily the entire insert of
          the specified clone. It may be shorter because we only sequence
          overlapping sections once, or longer because we arrange for a small
          overlap between neighbouring submissions.
          This sequence is the entire insert of clone F53F1. The true right
          end of clone M04G12 is at 21759 in this sequence. The start of this
          sequence (1..101) overlaps with the end of sequence Z81103.
          The end of this sequence (39379..39478) overlaps with the start of
          sequence AL021448.
          Location/Qualifiers
          1..39478
          /organism="Caenorhabditis elegans"
          /db_xref="taxon:6239"
          /chromosome="V"
          /clone="F53F1"
          complement(14923..16486)
          /gene="F53F1.1"
          complement(join(14923..15347,15663..15774,15835..15984,
          16033..16271,16318..16486))
          /gene="F53F1.1"
          /note="predicted using GeneFinder; similar to cuticlin"
          /codon_start=1

```

```

/protelin_id="CAB03124.1"
/db_xref="GI:3877456"
/translation="MDNKEIRRLLEDYEPKGFLLSESSKXDCHELDMLVLQLLATL
AVNPVPIONSLGCVQVECDSRITISQIKTEKPEVGIYFKDFASEEVCASRGGRIS
APFIEIGLGLRORVRLNPKGLAVRTITISFPYFIVKDRYINLCIGYRESQVIV
ANNISDEITISTISYNNVLMPTCTYQLSGCPREPERFELICQYVHOKCNDKDM
IILIKESFCMVVHTCSVDGREGTSFLIDNSGSDIKFLISLETGNNLAGGEAHV
YKPADRALFPOCISITVKEPQECVPRPICEDVEGGGAVGPPPYGMSKMLNVITM
OKNSFLTFKSPYMYVGVITNYHNKAIOKL"
/length=17396
/genes="F53F1.2"
join(17396..17599,17683..17926,17977..18146,18197..18286,
18339..18524)
/genes="F53F1.2"
/notes="similar to aldehyde reductase; cDNA EST yk47348.3
comes from this gene; cDNA EST yk47348.5 comes from this
gene"
/codon_start=1
/protelin_id="CAB03127.1"
/db_xref="GI:3877459"
/translation="MSDSFANVPQSGQKLSNGYNIPEVFGTYKVTGENYPAIDAL
TAGYRMPDZAKYVLTNEKEGALKILPKHGLSDVFLSKFEPEKNGRCRGFV
EESLQSIQPTIDYIMVLYVHPKPNDSDDVNNAYRKIAYEVLLEAKAGKRGVGS
NYETVHEEKTAKVPRPCANOLEYHPHARIPLOKCKREKNIEFOAFSSLARPERL
IEDPVVVELKKNHNSVPLVLLAMALRNQNGIVPKSVTISRIVENKVIDIALTPEDI
QSLTALDVGHHYIRCTGWLK"
/length=18714
/genes="F53F1.3"
join(18714..18887,18935..19056,19113..19302,19353..19448,
19919..20008,20072..20251)
/genes="F53F1.3"
/notes="similar to aldehyde reductase"
/codon_start=1
/protelin_id="CAB03128.1"
/db_xref="GI:3877460"
/translation="MSIKLNTGYDCPLIGCTYKTIQDQVLPVLDAAALTYGRLEDTA
KYVNNKEIGDALEILPKHNLKREDFITTKMHPNTVENVKRLVDSLSLTKTSYD
MGLIHPKSFQYDQDPMKTRLATNWDNMECKNAGKIRSVGSFEIHLRELKX
GKPPCCSNQVEYHPFTRELKNYKSCSEGIFFQAFSLARNHETLSSILIRLAEK
YHVPTVTLSSMATSQKVGITIPKSTNERLAQNKLVLLEEEYKVCNINLDXHYVR
TNGCIVL"
/length=21582
/genes="F53F1.4"
join(21582..21670,21781..22009)
/genes="F53F1.4"
/notes="predicted using GeneFinder; similar to cuticlin;
cDNA EST EMBL:D72018 comes from this gene; cDNA EST
EMBL:D7479 comes from this gene; cDNA EST EMBL:C11739
comes from this gene; cDNA EST EMBL:C12451 comes from this
gene; cDNA EST EMBL:C10189 comes from this gene; cDNA EST
EMBL:C10682 comes from this gene; cDNA EST yk506b3.3 comes
from this gene; cDNA EST yk437e9.3 comes from this gene;
cDNA EST yk437e9.5 comes from this gene; cDNA EST
yk311h10.3 comes from this gene; cDNA EST yk311h10.5 comes
from this gene; cDNA EST yk341f6.3 comes from this gene;
cDNA EST yk341f6.5 comes from this gene; cDNA EST
yk300g2.3 comes from this gene; cDNA EST yk300g2.5 comes
from this gene; cDNA EST yk234b3.3 comes from this gene;
cDNA EST yk234b3.5 comes from this gene; cDNA EST
yk191d9.3 comes from this gene; cDNA EST yk191d9.5 comes
from this gene; cDNA EST yk677d4.3 comes from this gene"
/codon_start=1
/protelin_id="CAB03125.1"
/db_xref="GI:3877457"
/translation="MOKIVIFFAIAVNSQAFLLPSSGGGGCGGACPPPPPPCCGGCS
LPLQLPRFELPLSLPSFGGCGGAPCGAAPPAYVADPPAAGYAAPPAAGYATA
GRK"
/length=25572
/genes="F53F1.5"
join(25572..25660,25850..26408)
/genes="F53F1.5"

```

```

/notes="predicted using GeneFinder; similar to cuticlin;
cDNA EST EMBL:T01107 comes from this gene; cDNA EST
CEESN60F comes from this gene; cDNA EST EMBL:D70966 comes
from this gene; cDNA EST EMBL:D71580 comes from this gene;
cDNA EST EMBL:D72120 comes from this gene; cDNA EST
EMBL:D72428 comes from this gene; cDNA EST EMBL:D73526
comes from this gene; cDNA EST EMBL:D74233 comes from this
gene; cDNA EST EMBL:D74402 comes from this gene; cDNA EST
EMBL:D74922 comes from this gene; cDNA EST EMBL:D75307
comes from this gene; cDNA EST EMBL:D76322 comes from this
gene; cDNA EST EMBL:C12136 comes from this gene; cDNA EST
EMBL:D69807 comes from this gene; cDNA EST yk505b11.3
comes from this gene; cDNA EST yk492b11.3 comes from this
gene; cDNA EST yk492b11.5 comes from this gene; cDNA EST
yk480h10.3 comes from this gene; cDNA EST yk480h10.5 comes
from this gene; cDNA EST yk445a7.3 comes from this gene;
cDNA EST yk445a7.5 comes from this gene; cDNA EST
yk440b8.3 comes from this gene; cDNA EST yk440b8.5 comes
from this gene; cDNA EST yk377b11.3 comes from this gene;
cDNA EST yk377b11.5 comes from this gene; cDNA EST
yk473g12.3 comes from this gene; cDNA EST yk473g12.5 comes
from this gene; cDNA EST yk459g4.3 comes from this gene;
cDNA EST yk459g4.5 comes from this gene; cDNA EST
yk482d10.3 comes from this gene; cDNA EST yk482d10.5 comes
from this gene; cDNA EST yk455c6.3 comes from this gene;
cDNA EST yk455c6.5 comes from this gene; cDNA EST
yk417a12.3 comes from this gene; cDNA EST yk417a12.5 comes
from this gene; cDNA EST yk414h1.3 comes from this gene;
cDNA EST yk401f4.3 comes from this gene; cDNA EST
yk401f4.5 comes from this gene; cDNA EST yk401e5.3 comes
from this gene; cDNA EST yk401e5.5 comes from this gene;
cDNA EST yk399e6.3 comes from this gene; cDNA EST
yk399e6.5 comes from this gene; cDNA EST yk392h1.3 comes
from this gene; cDNA EST yk392h1.5 comes from this gene;
cDNA EST yk391b10.3 comes from this gene; cDNA EST
yk391b10.5 comes from this gene; cDNA EST yk342c6.3 comes
from this gene; cDNA EST yk342c6.5 comes from this gene;
cDNA EST yk345c4.3 comes from this gene; cDNA EST
yk345c4.5 comes from this gene; cDNA EST yk307f5.3 comes
from this gene; cDNA EST yk307f5.5 comes from this gene;
cDNA EST yk258a5.3 comes from this gene; cDNA EST
EMBL:C09046 comes from this gene; cDNA EST yk306e6.5 comes
from this gene; cDNA EST yk314h1.5 comes from this gene;
cDNA EST yk331e10.5 comes from this gene; cDNA EST
yk343g4.5 comes from this gene; cDNA EST yk375f1.5 comes
from this gene; cDNA EST yk404c1.5 comes from this gene;
cDNA EST yk447c9.5 comes from this gene; cDNA EST
yk447e5.5 comes from this gene; cDNA EST yk488e6.5 comes
from this gene; cDNA EST yk382a5.3 comes from this gene;
cDNA EST yk306h10.3 comes from this gene; cDNA EST
yk507f12.3 comes from this gene; cDNA EST yk518b6.3 comes
from this gene; cDNA EST yk530f12.3 comes from this gene;
cDNA EST yk536e1.3 comes from this gene; cDNA EST
yk537c8.3 comes from this gene; cDNA EST yk540e5.3 comes
from this gene; cDNA EST yk545b10.3 comes from this gene;
cDNA EST yk557b11.3 comes from this gene; cDNA EST
yk562h11.3 comes from this gene; cDNA EST yk564d11.3 comes
from this gene; cDNA EST yk566f10.3 comes from this gene;
cDNA EST yk596a10.3 comes from this gene; cDNA EST
yk596e11.3 comes from this gene; cDNA EST yk614f1.3 comes
from this gene; cDNA EST yk618d8.3 comes from this gene;
cDNA EST yk620c8.3 comes from this gene; cDNA EST
yk620d5.3 comes from this gene; cDNA EST yk622e4.3 comes
from this gene"

```

Query Match 9.3%; Score 107.4; DB 34; Length 39478;
Best Local Similarity 61.7%; Pred. No. 2.1e-15;
Matches 171; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

```

Qy 329 gctcgaatccctgtcgtgaacatgtgtcgaacacatcgtatctgtcttcta 388
      ||||| |||| | ||| ||||| || ||||| || |||||
Db 15074 GCTCAGCGCGCTTGTACATCTTCACATATTGCTCAGCATTCCTGTCGCTTTG 15133
Qy 389 atgtaatactgatactgcatgatagaanaagctgtgatacgtccgacatattgtatacg 448

```



```

gene
CDS
    join(25834..25873,27277..27512,27582..27721,27768..28110,
    28158..28410,28491..28811,29092..29213)
    /gene="E04D5.3"
    /note="similar to cuticlin like protein"
    /codon_start=1
    /protein_id="CAA91280.1"
    /db_xref="GI:3875452"
    /db_xref="SPTREMBL:O19053"
    /translation="MPSRLAMKELKGGSEVVCETASISLFTKTRNSFGKVKY
    VSRPSCTMGDGTGTHREYRHMDCGRROREINGVYSITVLSFISITITDRAY
    RVSCFYEGTKKYNHNDISALTTOLESTOLPCRYETLNEGGSPVYARIGDQV
    YKMTCAELEENYCMKVHSCIVYDGGSPVVIDANGSVGVILQNLVTSDLTA
    GKLAPEKFKAGLYFNCOIQLIKDYNCGSTPOQCPSTOVVEPAKTEIAP
    YPDSHESGYPTRPAPVPSASRPITPTQAPASPSPPAPPGADIDNGYEPOPI
    YIAETPENAVDGIIVGENDTEQPTTSAAETEDGYSLIKRVVESTEOINASKKRP
    VTWGDIDLPERGILIVGELMEDEGTNAGDHGATRLRARNRSQEKTCSTSRMYT
    LILLCILFATTVVFIYIVOKROIILAQTAFFKP"
    29675..31348
    /gene="E04D5.4"
    join(29675..29840,31127..31214,31267..31348)
    /gene="E04D5.4"
    /codon_start=1
    /protein_id="CAA91281.1"
    /db_xref="GI:3875453"
    /db_xref="SPTREMBL:Q19054"
    /translation="MMSINLVHIIILVTFVSHVVTAVTKATGETTVAGODLGD
    VSSPFYETTTASTCADPNTDCTQYIFLCSNKKYPTLLQPFCAKTCGFCGSSSTAP
    VACVDNSTO"
BASE COUNT    10458 a    5592 c    5636 g    9850 t
ORIGIN
Query Match      7.2%: Score 83.8: DB 34: Length 31536:
Best Local Similarity 56.1%: Pred. No. 7.8e-10:
Matches 179; Conservative 0; Mismatches 137; Indels 3; Gaps 1:
QY 367 acattcgctatttgcttcttaacgtaatactgactgcatgataaagaactgtga 426
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 28101 ACATCCATAATTTACGCTTGTATTGAATTTGACAAATTTGAATTAACCCAGC 28042
QY 427 tgcattcgcatattgatactgactgctcttgccagcgcaataatcgttgatattc 486
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 28041 TTTGTCAGCAAAATTTGAAGACTGCTGCAGATTTCACAGCTGCAAGTCAGGCTACTC 27982
QY 487 caattatttagcaaatattatcaagaagcacatcat---cagcatttagaatttcac 543
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27981 AAGATTCTGAAGATCAACCCCATCTACAGAACATTCATGGCATCAATGACAGTTTACGG 27922
QY 544 agtacacgcttaccatcatcgacaagaacgaaatgacacgcgcagaaagtataaac 603
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27921 TGGTCACCTTGTGTCATCAATTAACCGTACGATGAACCTTTCATACAGTACATTTCTC 27862
QY 604 ggttcagaatcgacatgcatattataaactgctgctgccaatgataagcaaatgaac 663
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27861 AAGTTTCAGCAACGCAAGTTCATTGTGATAGACTTGCATTCGAGTGCAGCTACTTAAT 27802
QY 664 tgggtgacgcggttggtcga 682
    ||||| ||||| ||||| |||||
Db 27801 AGGTGATCCACCGGCTTCA 27783

```

```

RESULT 11
CEP53B6      32412 bp    DNA      INV      02-SEP-1999
DEFINITION  Caenorhabditis elegans cosmid F53B6, complete sequence.
ACCESSION   Z81086

```

```

VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

Z81086.1 GI:1627952
 HFG; Cuticulin; Human platelet tetraspan antigen like; Initiation factor associated protein; Thrombospondin like.
 Caenorhabditis elegans.
 Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditioidea; Rhabditidae; Pelodermidae; Caenorhabditis.
 1 (bases 1 to 32412)
 Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M., Bonfield,J., Burton,J., Connell,M., Copey,T., Cooper,J., Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A., Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M., Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N., Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B., O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A., Saunders,D., Shownkeen,R., Smaildon,N., Smith,A., Sonhammer,E., Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M., Vaughan,K., Waterston,R., Watson,A., Weinstock,L., Wilkison-Sporst,J. and Woldman,P.
 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
 Nature 368 (6466), 32-38 (1994)
 94150718
 2 (bases 1 to 32412)
 White,S.
 Direct Submission
 Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail: jess@sanger.ac.uk or rwenematode.wustl.edu
 Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information.
 For a graphical representation of this sequence and its analysis see:-
<http://webc.sanger.ac.uk/cgi-bin/display?db=wormacc&class=Sequence&object=F53B6>
 Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
 IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
 IMPORTANT: This sequence is not the entire insert of clone F53B6. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
 The true left end of clone F53B6 is at 1 in this sequence. The true right end of clone F53B6 is at 5116 in this sequence.
 The true left end of clone F32H2 is at 32306 in this sequence. The true right end of clone F08G11 is at 7984 in this sequence. The start of this sequence (1..105) overlaps with the end of sequence Z80220.
 The end of this sequence (32306..32412) overlaps with the start of sequence Z81523.
 Location/Qualifiers
 1..32412
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /chromosome="I"
 /clone="F53B6"
 1480..3649
 /gene="F53B6.1"
 join(1480..1638,1707..1946,2437..2628,2969..3154,3205..3308,3430..3649)
 /gene="F53B6.1"
 /note="predicted using GeneFinder; Similarity to Human platelet-endothelial tetraspan antigen (SWPFR3_HUMAN)"
 /codon_start=1
 /protein_id="CAB03120.1"
 /db_xref="GI:3877452"

```

FEATURES
    source

```

```

gene
CDS

```

```
gene
/translacion="MKSALGEEGRKKKQNLKRDPSKLANPLSEFSLYREDPITGLKN
RKIKNVQODSRLEFSLLEKYIFHSLISNLITLYLALHJOTRVPEPKSEPHNTLKF
OLKMGALDSAYGARGRLIKFSYIVTALISLIFISICIGYIMLLARSOYALVSS
LYVYGRILVILISLITNLICFYAIFKEMKRCVTSOAVASIVIAVLLIIGCGIGIN
PRDUTHTYPLNLKMLTSLRELTYCHDMKGTESMDALQSFCVYNGINDMAIKMT
SKWMHQAEPKLLIPESCCLPSEIERCSNFEODAPPYITSTCEPLONDLHVMN
VASWLCITMAIVQYIFVSLKTIKLNFLILTKT"
complement(4270..15311)
/gene="F53B6.2"
CDS
complement(join(4270..4339,4493..4547,5071..5295,
5368..5662,5891..6148,6204..6752,6818..7287,7546..7676,
8104..8184,12576..12775,13246..13384,13931..14132,
14183..14294,14341..14457,14994..15155,15198..15311))
/gene="F53B6.2"
/notes="predicted using GeneFinder: similar to
thrombospondin like: cDNA EST EMBL:Z14404 comes from this
gene; cDNA EST EMBL:C12571 comes from this gene; cDNA EST
yk195c9.3 comes from this gene; cDNA EST yk195c9.5 comes
from this gene; cDNA EST yk286g7.3 comes from this gene;
cDNA EST yk286g7.5 comes from this gene; cDNA EST
yk333e3.5 comes from this gene; cDNA EST yk353h1.3 comes
from this gene"
/codon_start=1
/protein_id="CAB03121.1"
/db_xref="GI:4008366"
/translacion="MKCSYTVVFLFYLILASPHVDALSNMAASPMSSCTKCGGYS
ROLRCLTSCSGSEVSRKVCQAKCTESKSLARDTICGGEIYISRGCEVQVCSRT
GANLRLMYDQSTPCQATSRVAVCSKSCQIYGCGLISSRFPACVCGRGRTCN
GKFLMKVSEETACASNCDDIVDWSGAGRSIASTSOPITVCVNAITRIVERKICAK
LRPRVEARPCPMLICPSRMMAAADTECVPHCGEGRREYVCQVTAHNVHVDTC
ENGTRPAEENCVCSTSCGRWAGKMSKCTASCGGVRRHVACVGGSDCEGGPRDE
TTCYAGIJCISATNSLDMNDRAVLDGNTGFSMDHMOAPRLVAGEMSTKSCSG
VMSSTVECVANPNISSAPIKLPMSECCODDEPKLFESCEVSRCPLOESKISEEAY
QMRGDMTQCSASGLCGKAKALCTIOYTSKQMSOCARRRRPREKSPRCNHPRP
PFLITSKYSDCSMSCSGSGTARRSKCAQYTSKTDGAHAHYLRDRCHFRKPTCTAC
NVACPRATWSSLNKRHNKIKLNLKTAQWTECSRSDSGERRQVWCETLDSGKTQ
RRPDECDANTPQTEVCSFGSCSRBELSNRVEEDNAEQKLTLLIGVALTYQGT
SIRIKCPACKEDKKIYWKNGKIKELNDAHIKVANGNLRFHARMEDAGYEEFTOR
LOGVTLNFKRDPSPASRVDPAPKPOIIPSTNRQROYKSCVLEKREOASVYHKMNVLI
IEALTPLNDEKAREOLARKYGNELVAMWDIGHMSECRKQKCHVAGYQARGISCVYTH
GEIRNVNDSITCESLASYRPEPTRCHREDCRWEASQSESSQRCVSSMLAQKRVN
TCRTNGTSDVIDIQCDITNRPATIMDCPNQCKAEMRTSDMGSCSSCGGGVQRL
SCWVSISSGRPAGRNCEQMRPHSARACVADPELPMPMTASATYQRDASCODQRFCD
IRLFHSCDSLSEVROKCCSTCTFEVERKKF"
complement(119892..20283)
/gene="F53B6.8"
CDS
complement(join(19892..20022,20241..20283))
/gene="F53B6.8"
/notes="F53B6.8"
/codon_start=1
/protein_id="CAB54255.1"
/db_xref="GI:5824510"
/translacion="MDFGILLVFLMAVAGTFAGISVSFSSHEPRHGLLGGGFNPY
GGYQSNYWGKRR"
complement(20516..21818)
/gene="F53B6.4"
CDS
complement(join(20516..20602,20648..20953,20999..21424,
21609..21818))
/gene="F53B6.4"
/notes="F53B6.4"
/notes="predicted using GeneFinder: cDNA EST yk193h3.3
comes from this gene; cDNA EST yk193h3.5 comes from this
gene; cDNA EST yk411f5.3 comes from this gene; cDNA EST
yk411f5.5 comes from this gene"
/codon_start=1
/protein_id="CAB03117.1"
/db_xref="GI:3877449"
/translacion="MIRISRGQPRFRKSKCDSDINKNOFEYLKEEARVCSIAITRML
VPLAIVTSLPMVNAIAFCARKNRTYHAKNNKNSKSKASKSTRGASKSGSKSR
SSKASHKRSKSKGTSKSGSKSGSKSGSKSKSKSKVKATITSGSOVSTVSA
TGVSDKOSNSKSKSKSKSKSRKRNRLDADAKKSKSKSKKVALLPDQOTTS
OYGHSLAEVNSTHKSKEMNYPAPKLOYOTLIGCJNOIELKNTSERKAYXIKGCDNL
YRNPVYGFAPRPSVAKIDVRLNGEOKTDLVLTIRANANGSTNPHAFANOAEHREH
NVVPLVAA"
```

```
gene
23555..24160
/gene="F53B6.5"
CDS
23555..24160
/gene="F53B6.5"
/notes="Similarity to Rat Initiation factor 2 associated
glycoprotein (Sw:12a6_RAT)"
/codon_start=1
/protein_id="CAB03119.1"
/db_xref="GI:3877451"
/translacion="MILSDNPLDQTKKEPDQASQKTESCEFHGNSNEVKPIGSLN
GQSLAQGRILHTEKTVLPVKGGEQARMENITVLAETFGTGKGFPHDMETSHYKMF
ELDAEKIPLRLQSKGLNLIDKNFATLACRCVMDIRLGETKXYLMALKGCSYSTASR
MLTLTLEFTDRMMAMVACILIQSPAKIIGKRANLLNLI"
26258..27459
/gene="F53B6.7"
CDS
join(26258..26342,26501..26802,26846..26926,26974..27107,
27153..27288,27334..27459)
/gene="F53B6.7"
/notes="predicted using GeneFinder"
/codon_start=1
/protein_id="CAB03118.1"
/db_xref="GI:3877450"
/translacion="MSRESVSLAQLVERETFLNKGSSPLVSHAESMGLIPPNWR
VCYRGRGMEYHYPDMNPKLVPRPLHYKDPDLPERGOIVSETYRGLINAGIRPF
EVCSPDMKSVOTAPFLIKGLSYTTINIDPALSYRMLPTNFOELSPRAEFM
GYPINIOYLSOGFIRAEINEDYNLRIOAFKKNIKILKQYVVISDNVMDLTRRE
HVEVVDILQICIKKPTQGMNFIILSKKEADIMSPILPLKSLYKPFVWTVDPLQ
SVNHPRLNQN"
26276..26348
/gene="F53B6.11"
/notes="Tyr Lys K-tRNA; predicted using tRNAscan-se-1.11;
predicted using prediction: similar to tRNA-Lys"
26276..26348
/gene="F53B6.11"
complement(27666..30589)
/gene="F53B6.6"
CDS
complement(join(27666..27862,27919..27959,28076..28770,
28905..29001,29330..29417,29467..29606,29648..29915,
30107..30301,30478..30589))
/gene="F53B6.6"
/notes="predicted using GeneFinder: similar to cuticulin"
/codon_start=1
/protein_id="CAB03116.1"
/db_xref="GI:3877448"
/translacion="MNLSTPLVTYVTFYVWLEKYVLHFKQYVVALRFESYQKNP
PILLLIYNRFQDITLILSTSSCFETQNGVCKPEVFCGIDTIRKVNTEHPN
GRIYVDESKOHQOVSHADASHSPQEFITPIDACNNRRORTLHPROISFPMITSE
HPFEVGTMDRAFSIRCFLESIRKGLNEIDVGTALPDHVOEYSLPVCAHYLMDGIEG
HVLRFAGOVGVTVHRCDDODASHVYGLILHSCYADDHGKRFELVDRCSTDPFL
PQIEYHGAISATVNAHVFRYKADKQVLYFCYVQOLCKKHGCGCEGITPPQSGSHSII
```

Query Match 6.48; Score 74.8; DB 34; Length 32412;
Best Local Similarity 59.3%; Pred. No. 1e-07;
Matches 127; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

RESULT 12
LOCUS CER07E3 39736 bp DNA INV 02-SEP-1999
DEFINITION Caenorhabditis elegans cosmid R07E3, complete sequence.
ACCESSION Z49207
VERSION Z49207.1 GI:1067021
KEYWORDS HTG; ADP/ATP carrier protein; Cuticulin; Cysteine proteinase.
SOURCE Caenorhabditis elegans.
ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida; Rhabditidae; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 39736)
AUTHORS Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rikken, L., Koopra, A., Saunders, D., Showkeen, R., Smaildon, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J. and Wohlman, P.
TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2 (bases 1 to 39736)
AUTHORS Cottage, A.
TITLE Direct Submission
COMMENT Submitted (05-MAY-1995) Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or twenematode.wustl.edu
On Nov 21, 1995 this sequence version replaced gi:798823.
Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information.
For a graphical representation of this sequence and its analysis see: "
http://webcace.sanger.ac.uk/cgi-bin/display?db=wormcace&class=Sequence&object=R07E3
Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones.
EXCEPTIONS are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
This sequence is the entire insert of clone R07E3. The true right end of clone F41E7 is at 36774 in this sequence. The start of this sequence (1..115) overlaps with the end of sequence Z68106.
The end of this sequence (39633..39736) overlaps with the start of sequence AL032624.
FEATURES
Source
Location/Qualifiers
1..39736
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="X"
/clone="R07E3"
5080..8873
/gene="R07E3.6"
join(5080..5246,5409..5769,5826..5942,6381..6905,6955..7075,7125..7448,7495..7787,7831..7980,8037..8174,8222..8405,8470..8873)
/gene="R07E3.6"
/note="Serine and threonine rich protein, Carboxyl terminus is glycine rich; cDNA EST EMBL:CO8296 comes from this gene; cDNA EST YK259b4.5 comes from this gene; cDNA EST YK559c11.3 comes from this gene; cDNA EST YK574b3.3 comes from this gene; cDNA EST YK597h8.3 comes from this gene"
gene
/codon_start=1

/protein_id="CAA89071.1"
/db_xref="GI:3878959"
/db_xref="SPTREMBL:Q21811"
/translation="MRVLCFTLVYVAARSAQFEFSMGNRENYVOFGKADATSSMDA LKTIAMRNPNKYNAPSRSSTIIPPTFEKAEDIDMKLRGKGDSTINKMILLPLRV KGQNPFTTSTTTSTTTTQKPTTSTTTSTTPTPTTTTATTTVLTTEELPK QOLAGRIQVOPDMSGEPSABDGIKKTKTAECVDDTDNNVSOADAKPOSTHNGK EHVSTTPESITENTTAEPASTASTKPTKLSVKIRSEDEKEBDKLDDEKPNYMG LEEIYKKEEELVERELAMSAVAIEPEKVVVTKTVPETSTTEPTTEATTTT EITTTTAKPTTTTOSTTTTQOLPTTTTTEPTTTTTPDIAENELKQATEVM KSLDEERIDAEKNVEYOKRDERMAAKRINNELKMKQAKRENNKHKEETTLAT EDPKTTTIVASTETSTVTDDTEATTTTATTTTATTTATTTKNSNIPESAEDEL EAKTTTKYEDGPIITTEPEYIKKRNALVIGRTLPQALEAINKAPISASSESAEP SSNETAAVSAASAAPVALPEPIAPDIALPOLPDIDISOEMRVASIVEDMRILG SILOSSASARRAAAVRYTEKTKPTNIRIDIMADGSENSIVGSGSLAREILNPI LKDKLIRKAMAKIEAAVRNNAVSNATGAPDLDSGPOQORVPATNSKLPYEF PPOGVGGLPPAPPPPTPPRPAPOVLPVOTQOTAPKPTLPPTTLPVLTYSKPTPKD EGYEDGRGVKTTTFHGDSCVMITGGSHSPTEDYILMPESESTGLAAVAPAPPPPTO ATAAVTPAPKVPAPSPFPGGGGGGGGNGKNGFGGSGSPVPFPEVERKNSK TDFSEFFSNRRKL"
/complement(11510..14654)
/gene="R07E3.5"
/complement(join(11510..11773,12101..12476,12528..12827,13402..13542,13587..13696,13745..13863,13912..14026,14541..14654))
/gene="R07E3.5"
/note="cDNA EST YK278a10.3 comes from this gene; cDNA EST YK278a10.5 comes from this gene; cDNA EST YK486f8.5 comes from this gene; cDNA EST YK603d8.3 comes from this gene"
gene
/codon_start=1
/protein_id="CAA89072.1"
/db_xref="GI:3878960"
/db_xref="SPTREMBL:Q21812"
/translation="MRLVFWFLIILLFLAVTGTCGLREYVNTLYIFPEKATLN EYEEBESVEDNEPDEFOHKKPPRSSASDGIINRKSIDIIDAKLETNYPEKOT TVSVLVDDTLDTITAGMEAVIEDOVTFNFSQAQLEPVLSTRTKTSFHFWMQILL IAGFMYRYVLPVPCRIALFGIALVYVTSIIGLVPAPKAKPLNRBCMLCMRTYS RAFSSVIRFHDKEKNRANKGICVANHTSPIDVMVLSDCNCAIMGOAKGFLGQTQ LRSSEHNIHFPERGEAGDRAKVMDRMREHVNENKLPJLIPEPGTINNTSVMEKGS FEIGTIVYPIVAKYTRTLDAPFNWSASGYGSLMSMTMAITODVYLPAMRGEN EDSISFAKRVRAITAKKGLIDLEMDGALKREKRSKTYTLOOKYFPRILRTTIN MLEKSSDILINMOSISEERDELKQIDEDDEEMIRKISSLKPKKRTGDHDAEH IKSD"
/complement(17917..19072)
/gene="R07E3.4"
/complement(join(17917..18025,18074..18331,18378..18555,18604..18751,18800..18877,18947..19072))
/gene="R07E3.4"
/note="similar to ADP/ATP carrier protein"
gene
/codon_start=1
/protein_id="CAA89069.1"
/db_xref="GI:3878957"
/db_xref="SPTREMBL:Q21809"
/translation="MASTSATLDVLETSKKCLAGSAANAISKTTTAPDRKVLVQLO ROSEFPAEAENGINICISIKIRLEDGAMALMGNGAGVARCPLNHTLNFAFDITRN LKNDVRESFGKFLAGTFPGSLGATLLFMVLPDEFARTLRLADLVKKDGRKYGW DCLKIKIRSECVASWYKGLSSALQFVISAIFGIPISITSVEDPSLNFACMAI AOISITTSWCWCPDITVRRSMMSQKQINQYSTRKCMKTLVKKQDINFGALIT NSLSTGALITTEYERSKYN"
/complement(21213..23259)
/gene="R07E3.3"
/complement(join(21213..21359,21654..21739,21787..21886,22011..22170,22220..22335,22428..22542,22637..22779,23254..23259))
/gene="R07E3.3"
/note="similar to cuticulin"
gene
/codon_start=1
/protein_id="CAA89068.1"
/db_xref="GI:3878956"
/db_xref="SPTREMBL:Q21808"
/translation="METSPPGLPLSTNVVAFNPEELTKNDRYKVOCFYEMERRIO KVIQISMPPTPMHSKQILNMPVCKYEVLDSGTPGPPVYFATGVQVHYHMCDDTHERNT FCMLVHSCFVDGNGORVOLINDKGCALDKYLLTNLEPTDLMAGREAHVYKADRON

MEDLINE 94150718
REFERENCE 2 (bases 1 to 36306)
AUTHORS McMurtry A
TITLE Direct Submission
JOURNAL Submitted (21-MAY-1996) Louis, MO 63110, USA. E-mail: jees@sanger.ac.uk or twenematode.wustl.edu
COMMENT On Nov 4, 1996 this sequence version replaced gi:1370039. Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information. For a graphical representation of this sequence and its analysis see:-
http://webace.sanger.ac.uk/cgi-bin/display?db=wormace&class=Sequence&object=M142

Current sequence finishing criteria for the C. elegans genome unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
This sequence is the entire insert of clone M142.
The true right end of clone C44B9 is at 5059 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence 273424.
The end of this sequence (36203..36306) overlaps with the start of sequence 299276.

FEATURES
Source Location/Qualifiers

1..36306
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="III"
/clone="M142"
3511..5795
/gene="unc-119"
join(3511..3570,4639..4750,4821..5201,5689..5795)
/note="unc-119"
/note="cDNA EST yk258a1.3 comes from this gene; cDNA EST yk258a1.5 comes from this gene; cDNA EST yk552a11.3 comes from this gene"
/codon_start=1
/protein_id="CAA97807.1"
/db_xref="GI:3878736"
/db_xref="SWISS-PROT:O10658"
/translation="MKAEQQQSIAPSGATFSPQMPRPPEVFOATTEAFLLKNOI TPNDVLAIPGTGSLCSPSNVNIETKQIRDLDEHVFELAKRENTENELQA QAESARVYRIFAPNFKLTKVGTATVEKGVDPITRMTERRHFFKRLKCFDEF GFCEPNSRNCEHIYEPFQISQQLMDIMNPNETRSDFYEVENKILVMHKADYSD A"
8540..12501
/gene="M142.2"
join(8540..8572,9335..9416,9468..9679,10373..10721,10772..10974,11033..11175,11871..12501)
/gene="M142.2"
/note="predicted using GeneFinder: similar to cuticulin"
/codon_start=1
/protein_id="CAA97806.1"
/db_xref="GI:3878735"
/db_xref="SWISS-PROT:O21540"
/translation="MRPIPYDISLSTFSLILICSNAPINDGLVSELIHECVTRK AVEVILLDASGSDGDFFKQLSFAMHLASRLNISEGSHMALIQAEPKLEFSLG QFNPTOLEMILQIREXOSGATNGOALRLTLELGGARGIKPVAIVLIDGOSLD VSEPSOLLRDADVVAIGVNLVNLHMQMGNPVAFVSEFEOIDRALIASLIM SMCKTEPRRPETITCGPDRIGVAKSKQPEGVNFVDHNDSECRGPEKFPDSRS IGLTVPSACVWHRIRSLNPKGITEVENSIVWFSLFTKTDQTVKQCFMEADKHV TVPLSVSMITTVFERQIYQMPQACAYTLARKAGDPPIVAFATLGESVYHMCIEVEGA KDRTFGMLVHSCYVDNGIGRVDILDSNGCGLDVLVLTSPDYDTSLRATPYVFX ADRLVFOCOITLCLKYDGSCEGITPRONCKKLGEGBHHHHHRRKRLVRLAD VGCTIDVFTSDVLEDEPACQVYTTYSFRSISTNCKIENS"
17427..23053
/gene="vab-7"

CDS
join(17427..17468,18628..18766,20758..20915,22487..22677,22840..23053)
/gene="vab-7"
/note="vab-7 even-skipped homologue"
/codon_start=1
/protein_id="CAA97809.1"
/db_xref="GI:3878738"
/db_xref="SWISS-PROT:O93899"
/translation="MOSPDIESLIGVNVPSLVEMVVASRASSFSPEEOHHDMGV VAAAAAARHHPIDNRDQGRYRTASRREGIGLEEFKENTVSKTSGELA AELNPEGTIVWQPNRMKDKRQKRVGGLAPLPLGPEPFLSQNSTKSPSSPHDSSKSK FGATGPGGAYGNNGSSTSPSAGSLPEPLGPEPFLSQNSTKSPSSPHDSSKSK NTSSDDSKSPVNSPSSSSPSPEYSTD"
28250..33043
/gene="M142.5"
join(28250..28366,28422..28499,28558..28991,30672..30966,31500..31763,32933..33043)
/gene="M142.5"
/note="predicted using GeneFinder: cDNA EST EMBL:701774 comes from this gene; cDNA EST yk470a9.3 comes from this gene; cDNA EST yk470a9.5 comes from this gene; cDNA EST yk476b6.5 comes from this gene; cDNA EST yk563c12.3 comes from this gene; cDNA EST yk662c9.3 comes from this gene; cDNA EST yk679e8.3 comes from this gene"
/codon_start=1
/protein_id="CAA97808.1"
/db_xref="GI:3878737"
/db_xref="SWISS-PROT:O21541"
/translation="MASCNLTAASNAIGACVIEITNGLSARGVISFDTTKVLYVD TREAMINPMIRINSEHLKNIKVLEALESEKFAARKTQDFQONNPNVNGRTTERL DTKGELTPNLMKSPAISIRGOQAVLQIKRIIPDTCWGEDIRVLGLVLRPYEVS DTKDAKATGPEDESRAESAVLOVKILSGPMETVYARAPDLTGAIYANSSMSDT DDIPLSADTILAAISMQAEQOEKIELOSGIIRKIDEDMLOSGFVREDTSRLTVE GVAALESSEARPARICGSSPTLYKPEHPEEKTKQIOITLFEFDDREKRTPEF VHYDKHPTDLPAELAKFDVITADPEPLAECLIKTAHSTRLLGKSDVKLCTGAI MEDYASRLMAHMRISFEFRHANNLANDPSCFANQTLTFC"
join(34039..34101,34448..34750,35530..36306,299276.1:105..698,299276.1:1364..1566,299276.1:1666..1772,299276.1:2495..2795,299276.1:3714..3928,299276.1:4056..4104,299276.1:5846..5958,299276.1:6478..6729,299276.1:7169..7338)
/gene="M142.6"
join(34039..34101,34448..34750,35530..36306,299276.1:105..698,299276.1:1364..1566,299276.1:1666..1772,299276.1:2495..2795,299276.1:3714..3928,299276.1:4056..4104,299276.1:5846..5958,299276.1:6478..6729,299276.1:7169..7338)
/gene="M142.6"
/note="similar to zinc finger, C3HC4 type (RING finger); cDNA EST EMBL:D67323 comes from this gene"
/codon_start=1
/protein_id="CAA97810.1"
/db_xref="GI:3878739"
/db_xref="SWISS-PROT:O45962"
/translation="MAPTGGQGGQVQVLCGSIQRHNEFPLVSLICGHVICKAE KPEQTKRCPHDWKTHSPSEVNNVALLSVIPRQCMILSGAVSAEARKVQSLI QIAKFEADSERGCVTSREISITLQKRYALLCQVREVDGKTLKMGKRGISEV MIEIISQSNTHVSQULMSVRAVRAQGFQAPADQVRLILITLLEEGEIAKRLV MYVQTLASDQVPOVSKTCVGHVOLLRASCFQVADQVRLILITLLEEGEIAKRLV EHDQIYQIAFESSLRIGPDMSKLLYADOSHRSKMSIIDKLOSKSYQOQVLELA LAGSQTMLVPARYFLTVYIPLCEFRAGIEHETSMKMGIDALHQIRLILKLCSD DLRMPKEERGVITLQAEVPGMGGGGSGGAGRIGLHNPYSDIDEGRSISPT NPKDINSINSPTPKQOROKRYQMGIPNNRGYSDDAPFIPSHQOQPPQPFNSQHL PQRGRGROGAPRPPQPPMPLIGDGMGAPMMAQTEVLTAGQVNVGAPQVAVM OSPTHLGGPVVMIPOQOAMPPOSPMIPVSGPBMGPMIPSPVQVAPMTATSP TGSYIYAASRPGRPPTTIQOSIGYVKKRSNPLIKYRKISFLNPFIDFLILKLLK EKKADILEFEKIKLSTDFKYPSSFSFGIMGMPMDRSGSGGVWGFTMLRESGADA EQLAKRYEILIKLQPSDEDDDPDGGIIGHVSAGRIGLHNPYSDIDEGRSISPT LPAIPFANMPTEETMTIGEMQVNRPAISLTVASSVLPNDPMNVNASSATVQAEEN RKIDPLFKRMKLTMEKVTSCFHVLLDQVYVYVSNLPNFASSRMRARRATIPQ PVIPMVQVPOVPIVPAENPNPNPVRPPRPGQPMVDSATIGLLPRLIIVAHQPN VVNSLDKTIQVYKERISEAQGNASAEANHLRMLRBAESOMAHLDPTTKNNCLRLA QOVDMLQQLHLNPTVEG"

BASE COUNT 11616 a 6626 c 6322 g 11742 t
 ORIGIN

Query Match 5.48; Score 62.6; DB 34; Length 36306;
 Best Local Similarity 52.08; Pred. No. 7.8e-05;
 Matches 196; Conservative 0; Mismatches 169; Indels 12; Gaps 2;

383 tcttaatggaatcaatcattgcatgagaaagcgtgacatccgcatattg 442
 12251 TATTTCAGACATAGTATTGACACGAAATGTACTACCGGGCGTCCCGACTTA 12192
 443 taaacgtg--agctctctggccagccatcaatcgttgatattccaaattatgac 499
 12191 AATACGTGTAGGTTCGTTCGGCAGACGACAGAGAGATGATCGCGAGTTAGAGA 12132
 500 aaatattataagacagcatccatcagcatttagaattccacagatcacccgttacca 559
 12131 AGCACCGCGTGTAGGCGCAACACATTCGAATCAGAAATATCCACTCTATACCGGAGCCG 12072
 560 tcatcagaacagcaggaatgagcaacccgcagagaatcaacggttcagaatc--- 616
 12071 TTAATCCATATGACAGAGATGACACTTCCGAAAGTCTCTTATACGACACCTCCACT 12012
 616 ----gaatgcatcattatgataaactggtcagccaatgataagcaattgactgtga 670
 12011 TCAATGCAATCCCATCTATGATATACACTTTCACCGAGAGATTGCAAAACGACATTTGGG 11952
 671 ccggttggtccaccatccaaattcaacgagcactgcatcgagacattgagtc 730
 11951 CCATCAGGGGGCGCTTTTTCGAGGGGTGAGACATTCGAGGCAATTTGATTTGCTCA 11892
 731 tgaagaagcagctgtgat 747
 11891 CGGAACAACCTGTTGAT 11875

RESULT 15
 CEZK265/36069 bp DNA INV 02-SEP-1999
 LOCUS Caenorhabditis elegans cosmid ZK265, complete sequence.
 DEFINITION 281143
 ACCESSION 281143.1 GI:1628350
 VERSION HMG; Collagen; Cuticulin; Homeobox protein; Human G9A protein like.
 KEYWORDS Caenorhabditis elegans.
 SOURCE Caenorhabditis elegans.
 ORGANISM Eukaryote; Metazoa; Nematoda; Secernentea; Rhabdilitia; Rhabdilitida; Rhabdilitidae; Rhabdilitoidea; Rhabdilitidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 36069)
 Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, J., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Latreille, P., Lingham, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Showkeen, R., Smaldon, N., Smith, A., Sonhammer, E., Stander, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstein, L., Walkinson-Sproat, J., and Wohlman, P.
 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans

TITLE
 JOURNAL Nature 368 (6466), 32-38 (1994)
 MEDLINE 94150718
 REFERENCE 2 (bases 1 to 36069)
 AUTHORS Dobson, R.
 TITLE Direct Submission
 JOURNAL Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail: jee@sanger.ac.uk or rwenematode.wustl.edu
 COMMENT Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information.
 For a graphical representation of this sequence and its analysis

FEATURES

source

gene

CDS

see:-
<http://webace.sanger.ac.uk/cgi-bin/display2db-wormaccdbclass=Sequence?object=ZK265>
 Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
 IMPORANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
 The true left end of clone ZK265 is at 1 in this sequence. The true right end of clone ZK265 is at 2412 in sequence Z75713.
 The true left end of clone T01G9 is at 35964 in this sequence. The true right end of clone T02E1 is at 4168 in this sequence. The start of this sequence (1..110) overlaps with the end of sequence Z81581.
 The end of this sequence (35964..36069) overlaps with the start of sequence Z75713.
 Location/Qualifiers
 1..36069
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /chromosome="I"
 /clone="ZK265"
 /complement(492..3554)
 /gene="ZK265.1"
 /complement(join(492..736,782..932,983..1189,1629..1758,1943..2076,2124..2289,2365..2762,3174..3313,3365..3446,3499..3554))
 /note="ZK265.1"
 /note="Predicted using GeneFinder; similarity to Human G9A protein (tr:G287865)"
 /codon_start=1
 /protein_id="CAB03514.1"
 /db_xref="GI:3881550"
 /db_xref="SPTREMBL:O94400"
 /translation="MAEVSVAIVRLKISALSADKAIPSAODEICEKEEPTNANFGVPHDDLNQSSSTIRPYGHPGLSLVPPFPNKSRYVANSRENLOKMNLCNOEIMFDTHTKNTILERKILDSOTPDSTYFLSHFPMKESSTRIKTCVHPAPFIDQECIRHMSHSVANSVSSASADLDORVHANLAKIAMLRYNKEFTISVNPANPFAEDSSCVVEYPGIKOCRAQRAASSTQDMRSRLPLHKAEDNAEIRFLKIGDMSNLRDDSMPELHACFEGHLEIVHEHLSNPOMAIANONGATLQYAVINENYLVETLTSASIDVNTONNNGEYPRIDYCGNHPAIQKILELQIFKSIIVDTYSIGSFSSRSPEATVGEVLIDARETOINREOMNCFALFYVSESSMLOLQDSDLDKLYKDKNTTRKLVGDIFETPRILKLRNAYATRMELTVYSIEAVIFINIFDIKRNORSYSDOH"

gene
 CDS
 5062..6428
 /gene="ZK265.3"
 join(5062..5393,5442..5628,5925..6428)
 /gene="ZK265.3"
 /note="predicted using GeneFinder"
 /codon_start=1
 /protein_id="CAB03515.1"
 /db_xref="GI:3881551"
 /db_xref="SPTREMBL:O94401"
 /translation="MPYAKEDSWAFOPIGSPPEPAPVRYVNPONNOVVALYWKKGKPIHGRAMNNGVYECSPFYKAKELKGLDGGOTIOILOYGDIYNSLGYWEMLPLKORHENNEGIRETVPGNSVPLAKLKGDLKGFIDNLNTEVALFNSAGTEYFEGGATANFMTIFRNLRPPTPGILKYVDLWYDLRQDNFNSNAVPADRALNTEGTEPMQVVALYWKGDVPEGRVYSSAGKTMHAFKKNNDENAGPEVSGMQLTVEASCMGLEIYKMPPLSEKSSGWTGTVHINSAPCILKDDKGLVLELNLDTTEKASAGYGGKEKIIMTGAAVAKLVLFKRLIA"
 /complement(6676..7870)
 /gene="ZK265.2"
 complement(join(6676..7754,7804..7870))
 /gene="ZK265.2"

```
/note="Predicted using GeneFinder; similar to collagen;
CDNA EST EMBL:D69143 comes from this gene; CDNA EST
EMBL:D65641 comes from this gene"
/codon_start=1
/protein_id="CAB03513.1"
/db_xref="GI:3881549"
/db_xref="SPTREMBL:Q94399"
/translation="MTTSGQLVAGTAVALACVAIPALFAFLVLIHNDINSFQSGVED
LAERFLAEDAMTMMVPREDAVOPEETFLKRRGQADGAGSAGSSGCGPOPK
DCAFPGRPCPNPGTGGDDGAGRAGNCGSDSTGGDRADFNKDKVCPAGPGRPN
GFPGPGDGGDGGGDKNGKDEPGDGGDEGTPGLPGPGEDEGPGGNGTGGGCG
PGVYAGRAGCGRGGDEPGENGODGCGRPGRAGDQPGHGGDGGSGVGEYGA
GADAAVCPGPRSAEMATGSSDSQSPASVEAPAPATKGYDSPAPAAKGYDAPAPTA
PHPPAPVAPPKLHDYESPAPVADHADPAAPQYKRRKVAFAHNV"
/complement(8777..13818)
/gene="ceh-8"
/complement(join(8777..8956,9538..9646,9702..9830,
9879..9976,12270..12598,13623..13818))
/gene="ceh-8"
/note="Predicted using GeneFinder; Similarity to Mouse
homeobox protein S8 (TR:G51366)"
/codon_start=1
/protein_id="CAB03519.1"
/db_xref="GI:3881555"
/db_xref="SWISS-PROT:Q94398"
/translation="MSNPTDNNSATSSITPNCERLMEVSTNLVIOFTYFKLYPNYF
QIRASFGAPPPLLENNKIKCFSAKAIQLPKLISFFSSAFSEFTYPPSDS
ESPESSPSKSSSSKRNKDIADKRRRRRTTETQLALEAFKMTJPPVYAR
ETLAAVQLPEVRYOVNOMRRAKFRROEKDDCGEELHSLKIDMPSMSKXTDT
PMLPRANTLSTNNNGISDEFEKTSKKYKYPFRAYGCPAANTHYSKGNVPHLN
FEDSDKSMKESQTPSTSSPFSIENHPPIYIIPNQSFAFNQIPMPFPIPIHF
EPOLTHSOENC"
/complement(16560..21861)
/gene="ZK265.5"
/complement(join(16560..16667,16744..16929,16977..17133,
17179..17280,17420..17619,17696..17765,17814..17848,
18123..18173,19898..20006,20052..20188,20236..20358,
20406..20562,20879..21047,21096..21217,21445..21489,
21572..21710,21816..21861))
/gene="ZK265.5"
/note="CDNA EST YK48195.5 comes from this gene; CDNA EST
YK50864.3 comes from this gene"
/codon_start=1
/protein_id="CAB03517.1"
/db_xref="GI:3881553"
/db_xref="SPTREMBL:Q94403"
/translation="MINATEIEIREDIHADLTVMAMPFRNSHOMPSIGVLAOVL
IISRTFLIGSTAMPFOKRLIKVOKLFTYRIEHRNORNIISLILMPELLCRY
LLAARFNLAIEIESIKNDECHHFGCSYAKMDESIPTAANRRHYMKFYITTPY
GILTERIPATLVLDYERKSRHMLFGLIEVAQNIFSCMAVVTTSGLTFOVLISGI
FSLPCSAVIALVEYFNOQLMTERDRTNNTLSIRYQLENIKTLKVMRREISL
ILIIAMGLISPLVILNEDILITVAVMYVSNPVLFTAIETIENRYKTF
NKARTTGMRIIDSRKDEMEKLELSDSRPSIIMSTRSSRADSTTKRPASNPSTP
NAALGIVATAIROLIFIDARKVALFYLAFTVLSPIESRIELDSYIYLVOKHVLNOY
GVKMGWTLIVYPTFIFMSKANNRRDOPIDVYCLRGVGTACWTFSYOPFFKVL
LTSWCDGRILTRAQCSKEGVTMPGDIOSGFLMITSILITEALAYRHYQVTD
AVHQMDDREHDLTRCIOYFVAMLEFLAHAFWEKQIIISVLYHIFIEELIGAAV
VCMFTVYRMLYPAGFLASPIRRIVGRK"
/18275..19128
/gene="ZK265.6"
/join(18275..18384,18431..18539,18656..18751,18874..19128)
/gene="ZK265.6"
/note="Predicted using GeneFinder"
/codon_start=1
/protein_id="CAB03516.1"
/db_xref="GI:3881552"
/db_xref="SPTREMBL:Q94402"
/translation="WARDVKRKGKPAYTNRNRKQYLKKDKNKKLSKSAVPIIKYTW
DEKTPRENVNDMGIAFNPNDAVPIAEHRKEIDAEPIDGSIVNPPKKEKVGKKK
EKOAAHIIISNLEQOVKEEERAGQDRFLFHRHRETELGYMLARHGDEFOAMTRDPK
NLMQYTPKMAKKIRTHKSEMKCLETA"
/27566..31619
/gene="ZK265.7"
```

```
CDs
join(27566..27635,27677..27835,28839..28913,29226..29303,
29467..29631,29755..29902,29982..30024,30072..30293,
30569..30709,30782..30854,30908..30989,31418..31619)
/gene="ZK265.7"
/codon_start=1
/protein_id="CAB03518.1"
/db_xref="GI:3881554"
/db_xref="SPTREMBL:Q94404"
/translation="MRGTSILPNPAPKMLEISRLPNFTSTFLIFPLHFKAKEM
SEICVEICLRVILDKNPVGRKLKRAKKNGYGLVEDCPTAAGCIRAKKHIE
MYELSRILYRNOILVCYVDEEYKSMTCMRKSGSIRMGCGRENCNDPERSRLYFAF
TEGNDVADKRISEMLKSEKDESSMKKFTTEPEPLETTTKKILITTPKRTTKTTT
TTSTSTSTTKENSFPLVDAKLPVLSITFYLKKEFSIKQNVFISNLYFRPTTA
TTTARVTPKSKTRIGGTTAHPLEPKDASKARVINVIDIRNHATSLWYQFVF
KGVLPNDISLDDTFEETRRKLEKEMEERLEMLADEEENSNDISDIDDDGI
TAEBORREERREYRMRRAERRLQEDTILRLIEEERIRADKDRMRREADDMYG
SDFENSTGSLVSVSAVILLIV"
/join(33403..33438,34288..34402,34791..34918,34995..35304,
35355..35799,35846..36069,275713.1:107..118,
275713.1:188..290,275713.1:343..430,275713.1:495..556)
```

```
Query Match 5.4% Score 62.2; DB 34; Length 36069;
Best Local Similarity 56.7%; Pred. No. 9.8e-05;
Matches 115; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 691 aattcataacgacatcgtgacatcgagacatcttgagtttgaaagcagttgatttc 750
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 35198 AATTGAATATGATGCATCCAGTATTCATGTCCTGCTTATTCGTGCGTGAATCAT 35139

QY 751 agatacccaatcgttgactaacgcttltacaacttccatgataagcattgaccg 810
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 35138 ACTAACACCCACATCAGCAGTCAGCCCTTATTCCTTCTTCAAAAAGCATTTCACATG 35079

QY 811 atatgcagatcaacttgggtacaaataatgataagaaacgaatgacacagttgtgt 870
||||| ||||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 35078 AATGCTTGATCACTTTAGTATGATGAATGAGGAAAGACAGACAGCTGATACC 35019

QY 871 tacaaaataaccagttgattca 893
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 35018 GTAGTACATTCACAGCTGATCCA 34996
```

Search completed: April 16, 2000, 04:58:10
Job time: 11299 sec

gene

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: April 16, 2000, 04:45:32 ; Search time 265.24 Seconds
(without alignments)
1095.133 Million cell updates/sec

Title: US-09-323-427-5
Perfect score: 1161
Sequence: 1 tgcctctgattgacgaa.....gcataagacacatcatc 1161

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues
Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	3.4	110000	1 V21209_06	Continuation (7 of 17) of DNA encoding a N-A
2	35.6	3.1	541	1 T83869	S. aureus coding s
3	35.4	3.0	348	1 X33788	S. aureus coding s
4	35.4	3.0	1182	1 X33781	S. aureus coding s
5	35.4	3.0	7563	1 V74344	P. falciparum gp19
6	34.8	3.0	4940	1 V35363	Sequence encoding
7	34.8	3.0	5760	1 M50530	Enterococcus faeca
8	34.8	3.0	8395	1 X13154	Enterococcus faeca
9	34.6	3.0	1479	1 X13342	Enterococcus faeca
10	34.6	3.0	2155	1 V23659	Human interleukin-
11	34.6	3.0	9062	1 V74396	Staphylococcus aur
12	34.6	3.0	21252	1 X13031	Enterococcus faeca
13	34.6	3.0	110000	1 X20248_02	Continuation (3 of 17) of Enterococcus faeca
14	34.2	2.9	3549	1 X13796	Enterococcus faeca
15	34.2	2.9	4865	1 T36122	S. mutans antigen
16	34	2.9	2004	1 T85356	Nephila clavipes s
17	34	2.9	6035	1 V74583	Staphylococcus aur
18	33.8	2.9	32768	1 X13037	Enterococcus faeca
19	33.6	2.9	110000	1 X20248_08	Continuation (9 of 17) of Bacillus protease
20	33.4	2.9	1867	1 Q40424	Enterococcus faeca
21	33.4	2.9	3327	1 X13239	Maize flury2 gene
22	33.4	2.9	10529	1 V09028	L.lactis DB1341 Ad
23	33	2.8	2088	1 V23689	Sequence of fowipo
24	33	2.8	3185	1 V23685	Staphylococcus aur
25	33	2.8	11225	1 V91695	Sequence of alpha-
26	32.8	2.8	838	1 V75138	Sequence of alpha-
27	32.8	2.8	1016	1 N93636	Sequence of alpha-
28	32.8	2.8	6978	1 N90096	Continuation (2 of 17) of Staphylococcus aur
29	32.6	2.8	110000	1 V21209_01	DNA sequence used
30	32.6	2.8	252	1 V77404	DNA encoding therm
31	32.6	2.8	891	1 V37169	Staphylococcus aur
32	32.4	2.8	4090	1 T86703	Staphylococcus aur
33	32.2	2.8	4418	1 V74765	Staphylococcus aur
34	32.2	2.8	110000	1 T58840_0	Mycoplasma genital

35	32.2	2.8	110000	1 X20248_07	Continuation (8 of 17) of Enterococcus faeca
36	32	2.8	716	1 X13803	DNA encoding one u
37	32	2.8	1709	1 T84145	DNA encoding a X-P
38	32	2.8	1709	1 V53488	N.clavipes draglin
39	32	2.8	2338	1 Q14183	Nephila clavipes s
40	32	2.8	2338	1 V23249	Staphylococcus aur
41	32	2.8	3775	1 V74549	Staphylococcus aur
42	32	2.8	3787	1 V74686	Staphylococcus aur
43	32	2.8	7159	1 V74613	Staphylococcus aur
44	31.8	2.7	357	1 V76836	Human IL-1ra BAC c
45	31.8	2.7	3972	1 X03040	

ALIGNMENTS

RESULT 1	V21209_06	Continuation (7 of 17) of V21209 from base 600001 (Methanococcus jannaschii circular
WP	Sequence split into 17 fragments	LOCUS V21209 Accession V21209
WP	Fragment Name	Begin End
WP	V21209_00	1 110000
WP	V21209_01	100001 210000
WP	V21209_02	200001 310000
WP	V21209_03	300001 410000
WP	V21209_04	400001 510000
WP	V21209_05	500001 610000
WP	V21209_06	600001 710000
WP	V21209_07	700001 810000
WP	V21209_08	800001 910000
WP	V21209_09	900001 1010000
WP	V21209_10	1000001 1110000
WP	V21209_11	1100001 1210000
WP	V21209_12	1200001 1310000
WP	V21209_13	1300001 1410000
WP	V21209_14	1400001 1510000
WP	V21209_15	1500001 1610000
WP	V21209_16	1600001 1664976

Query Match	3.4%	Score 40:	DB 1:	Length 110000:
Best Local Similarity	47.9%	Pred. No. 0.54:		
Matches 115:	Conservative	0:	Mismatches 125:	Indels 0:
Gaps 0:				
QY	256	tccagatcttcttgagaaacgaagtcgcacgtgcagcaggttgcgcgcacc	315	
DB	84567	TACAGAAATCATTAATAACTTTTAAAGAAATTTGCCGCTCAAAAGGAATATGACATTAC	84626	
QY	316	acctgtttaacagctccgaatccttggtctgaacatggtggtcgaacacatcgc	375	
DB	84627	ACATGAAAAACAGCTTCTTATTGTTTAAAGATTGACATGCTGTTCCGCAAAATACCA	84686	
QY	376	atttggtcttaatagttaatactatcgtgcattgatagaagaacgtgatcgcgc	435	
DB	84687	GATTGTGAGTTGAGAGGAACCTCGATTGCTCAAGTCACACAGCTCAATTTCAG	84746	
QY	436	atatgttaacgttgagctcttcgcgcacatlaaactgttgatattcaaatatt	495	
DB	84747	TCAGAAAGATTGGAAGAACTTAGAACATTAAATTAAGAGTTAAATTTGATTAATTAAT	84806	
RESULT 2	T83869/c	standard; DNA: 541 BP.		
ID	T83869			
DT	20-AUG-1998	(first entry)		
DE	DNA encoding a N-acetylglucosamine-6-phosphate deacetylase.			
KW	Staphylococcus aureus protein; ribozyme; antisease sequence; control;			
KW	Staphylococcal gene; regulatory element; bacterial gene expression;			
KW	vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;			
OS	toxic shock syndrome; ss.			
Key	Staphylococcus aureus.			
Location/Qualifiers				

FT CDS 20. .358
 FT /tag= a
 PD M09730070-A1.
 PD 21-AUG-1997.
 PD 19-FEB-1997; U02318.
 PR 20-FEB-1996; US-011888.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO,
 PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;
 DR WPI: 97-424969/39.
 DR P-PSDB: W27907.
 PT Novel polypeptide(s) from *Staphylococcus aureus* strain WCUH29 - used
 PT to isolate antimicrobial compounds, and in vaccines against *S.*
 PT aureus infection
 PS Claim 9: Pages 719-720: 989pp: English.
 CC The present sequence encodes a *Staphylococcus aureus* protein, that,
 CC based on homology with an *Escherichia coli* protein, is believed
 CC to be a N-acetylglucosamine-6-phosphate deacetylase. The present sequence
 CC was obtained from a library of clones of *S. aureus* WCUH 29 in *Escherichia*
 CC coli. The DNA sequence can be used in the construction of ribozymes and
 CC antisense sequences to control the expression of *Staphylococcal* genes.
 CC The DNA sequence is also useful as a source of regulatory elements for
 CC the control of bacterial gene expression. The encoded protein may be used
 CC to produce vaccines to enable a host to produce specific antibodies
 CC with antibacterial action. These vaccines and antibodies would protect
 CC a host against invasion by *S. aureus*, and conditions relating to
 CC *Staphylococcal* infection, e.g. *Staphylococcal* food poisoning, scaled
 CC skin syndrome, and toxic shock syndrome.
 SQ Sequence 541 BP; 180 A; 75 C; 107 G; 171 T;

Query Match 3.1%; Score 35.6; DB 1; Length 541;
 Best Local Similarity 52.5%; Pred. No. 0.83;
 Matches 74; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

OY 1006 aacatgcttcgcgaatgcatacgtatgataaattgattgttaagttggtccaca 1065
 Db 229 ACCATTTCGAACAGCGTGTGTCGANTGACAGTACTTTTTCGCCACCAATACATA 170
 OY 1066 ttcaatttcgtcaccatcgtcacaccatgtcaccggaaatcgaaatgaacaaatgcatt 1125
 Db 169 TTCTCCTTCAGGCAATACCTTTTGCACGCAATTCATCGTAATTAATAAANACGTTCAAT 110
 OY 1126 aagtgtagacagaagaacat 1146
 Db 109 ACCTTTCATACGTAAGCAAT 89

RESULT 3
 X33788/c
 ID X33788 standard; DNA: 348 BP.
 AC X33788;
 DT 25-JUN-1999 (first entry)
 DE *S. aureus* coding sequence SEQ ID NO. 45.
 KW *S. aureus* infection; diagnosis; therapy; central nervous system disorder;
 KW upper respiratory tract infection; otitis media; bacterial tracheitis;
 KW acute epiglottitis; thyroditis; empyema; lung abscess; splenic abscess;
 KW cardiac infection; infective endocarditis; secretory diarrhoea; ulcer;
 KW retroperitoneal abscess; cerebral abscess; blepharitis; conjunctivitis;
 KW keratitis; endophthalmitis; preseptal cellulitis; orbital cellulitis;
 KW dacryocystitis; epididymitis; intrarenal abscess; perinephric abscess;
 KW toxic shock syndrome; impetigo; folliculitis; cutaneous abscess;
 KW cellulitis; wound infection; bacterial myositis; septic arthritis;
 KW osteomyelitis; *Helicobacter pylori* infection; stomach cancer; gastritis;
 KW ss.
 OS *Staphylococcus aureus*.
 PN M09912557-A1.
 PD 18-MAR-1999.
 PD 14-SEP-1998; U18987.
 PR 12-SEP-1997; US-058710.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Burnham MKR, Lometto MA, Warren PV;
 PI WPI: 99-229138/19.
 DR

DR P-PSDB: Y05315.
 PT New isolated *Staphylococcus aureus* polynucleotides
 PS Claim 21: Page 90-91: 102pp: English.
 CC This sequence represents a *S. aureus* polynucleotide of the invention.
 CC The invention also relates to the polypeptides encoded by the
 CC *S. aureus* polynucleotides.
 CC The polypeptides can be used for the treatment or prevention of disease.
 CC The polypeptide or polynucleotide can also be used to diagnose diseases
 CC related to their expression. The polypeptides and vectors containing them
 CC can also be used in immunisation methods. The products can be used for
 CC treating infection, e.g. infections of the upper respiratory tract,
 CC (e.g. otitis media, bacterial tracheitis, acute epiglottitis,
 CC thyroditis), respiratory (e.g. empyema, lung abscess), cardiac
 CC (e.g. infective endocarditis), gastrointestinal (e.g. secretory
 CC diarrhoea, splenic abscess, retroperitoneal abscess), central nervous
 CC system (CNS) (e.g. cerebral abscess), eye (e.g. blepharitis,
 CC conjunctivitis, keratitis, endophthalmitis, preseptal and orbital
 CC cellulitis, dacryocystitis), kidney and urinary tract
 CC (e.g. epididymitis, intrarenal and perinephric abscess, toxic shock
 CC syndrome), skin (e.g. impetigo, folliculitis, cutaneous abscesses,
 CC cellulitis, wound infection, bacterial myositis), bone and joint
 CC (e.g. septic arthritis, osteomyelitis), or *Helicobacter pylori*
 CC infections, (e.g. causing stomach cancer, ulcers and gastritis). The
 CC products can also be used for treating in-dwelling devices and wounds.
 SQ Sequence 348 BP; 108 A; 55 C; 78 G; 107 T;

Query Match 3.0%; Score 35.4; DB 1; Length 348;
 Best Local Similarity 53.2%; Pred. No. 0.77;
 Matches 75; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

OY 1006 aacatgcttcgcgaatgcatacgtatgataaattgattgttaagttggtccaca 1065
 Db 222 ACCATTTCGAACAGCGTGTGTCGANTGACAGTACTTTTTCGCCACCAATACATA 163
 OY 1066 ttcaatttcgtcaccatcgtcacaccatgtcaccggaaatcgaaatgaacaaatgcatt 1125
 Db 162 TTCTCCTTCAGGCAATACCTTTTGCACGCAATTCATCGTAATTAATAAANACGTTCAAT 103
 OY 1126 aagtgtagacagaagaacat 1146
 Db 102 ACCTTTCATACGTAAGCAAT 82

RESULT 4
 X33781/c
 ID X33781 standard; DNA: 1182 BP.
 AC X33781;
 DT 25-JUN-1999 (first entry)
 DE *S. aureus* coding sequence SEQ ID NO. 8.
 KW *S. aureus* infection; diagnosis; therapy; central nervous system disorder;
 KW upper respiratory tract infection; otitis media; bacterial tracheitis;
 KW acute epiglottitis; thyroditis; empyema; lung abscess; splenic abscess;
 KW cardiac infection; infective endocarditis; secretory diarrhoea; ulcer;
 KW retroperitoneal abscess; cerebral abscess; blepharitis; conjunctivitis;
 KW keratitis; endophthalmitis; preseptal cellulitis; orbital cellulitis;
 KW dacryocystitis; epididymitis; intrarenal abscess; perinephric abscess;
 KW toxic shock syndrome; impetigo; folliculitis; cutaneous abscess;
 KW cellulitis; wound infection; bacterial myositis; septic arthritis;
 KW osteomyelitis; *Helicobacter pylori* infection; stomach cancer; gastritis;
 KW ss.
 OS *Staphylococcus aureus*.
 PN M09912557-A1.
 PD 18-MAR-1999.
 PD 14-SEP-1998; U18987.
 PR 12-SEP-1997; US-058710.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Burnham MKR, Lometto MA, Warren PV;
 PI WPI: 99-229138/19.
 DR P-PSDB: Y05308.
 PT New isolated *Staphylococcus aureus* polynucleotides
 PS Claim 20: Page 65-67: 102pp: English.
 CC This sequence represents a *S. aureus* polynucleotide of the invention.

CC The invention also relates to the polypeptides encoded by the
CC S. aureus polynucleotides.
CC The polypeptides can be used for the treatment or prevention of disease.
CC The polypeptide or polynucleotide can also be used to diagnose diseases
CC related to their expression. The polypeptides and vectors containing them
CC can also be used in immunisation methods. The products can be used for
CC treating infection, e.g. infections of the upper respiratory tract,
CC (e.g. otitis media, bacterial tracheitis, acute epiglottitis,
CC thyroditis), respiratory (e.g. empyema, lung abscess), cardiac
CC (e.g. infective endocarditis), gastrointestinal (e.g. secretory
CC diarrhoea, splenic abscess, retroperitoneal abscess), central nervous
CC system (CMS) (e.g. cerebral abscess), eye (e.g. blepharitis,
CC conjunctivitis, keratitis, endophthalmitis, preseptal and orbital
CC cellulitis, dacryocystitis), kidney and urinary tract
CC (e.g. epididymitis, intrarenal and perinephric abscess, toxic shock
CC syndrome), skin (e.g. impetigo, folliculitis, cutaneous abscesses,
CC cellulitis, wound infection, bacterial myositis), bone and joint
CC (e.g. septic arthritis, osteomyelitis), or Helicobacter pylori
CC infections, (e.g. causing stomach cancer, ulcers and gastritis). The
CC products can also be used for treating in-dwelling devices and wounds.
CC Sequence 1182 BP; 416 A; 174 C; 257 G; 335 T;

Query Match 3.0%; Score 35.4; DB 1; Length 1182;
Best Local Similarity 53.2%; Pred. No. 1.4;
Matches 75; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Oy 1006 aacatgctccgcgaatcgatcgatgataaattgattgtatgaattggtgcaca 1065
Db 945 ACCATTTCGAAGACCTGCTTTGGGATTGACACTTACTTTTCCACCCCAATTCATA 886
Oy 1066 ttcaattcttgctacactgcacacacattgcaacgcgaatcgataagaatgcatt 1125
Db 885 TTCTCCTTCAGGACATACCTTTTGGACGATTCATCGGTAAATTAATAAAGCTTCATT 826
Oy 1126 aagtgtagtacagaagaacatt 1146
Db 825 ACCTTTCATACGGTAAGCAAT 805

RESULT 5
V74344/C
ID V74344 standard; DNA: 7563 BP.
AC V74344;
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus entry SEQ ID #33.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT misc_feature 481..540
FT /*tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature 2281..2340
FT /*tag= b
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature 4081..4140
FT /*tag= c
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature 5881..5940
FT /*tag= d
FT /note= "these bases represent a line of missing text in

FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT

PN EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA,
DR WPI: 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1: Page 321-325; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 7563 BP; 2584 A; 1028 C; 1476 G; 2229 T;

Query Match 3.0%; Score 35.4; DB 1; Length 7563;
Best Local Similarity 53.2%; Pred. No. 3.2;
Matches 75; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Oy 1006 aacatgctccgcgaatcgatcgatgataaattgattgtatgaattggtgcaca 1065
Db 5616 ACCATTTCGAAGACCTGCTTTGGGATTGACACTTACTTTTCCACCCCAATTCATA 5557
Oy 1066 ttcaattcttgctacactgcacacacattgcaacgcgaatcgataagaatgcatt 1125
Db 5556 TTCTCCTTCAGGACATACCTTTTGGACGATTCATCGGTAAATTAATAAAGCTTCATT 5497
Oy 1126 aagtgtagtacagaagaacatt 1146
Db 5496 ACCTTTCATACGGTAAGCAAT 5476

RESULT 6
V35363
ID V35363 standard; DNA: 4940 BP.
AC V35363;
DT 23-SEP-1998 (first entry)
DE P. falciparum gp190 DNA.
KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
KW monoclonal antibody; passive immunisation; parasite; ss.
OS Plasmodium falciparum.
FH Key Location/Qualifiers
FT CDS 10..4929
FT /*tag= a
FT /product= gp190
FT
FT WO9814583-A2.
FT 09-APR-1998.
PD 02-OCT-1997; E05441.
PF 02-OCT-1996; DE-040817.
PR (BUSA/) BUJARD H.
PA (BUSA/) BUJARD H, Pan W, Tolle R;
PI Bujard H, Pan W, Tolle R;
DR WPI: 98-240088/21.


```
Db 4114 CCATCTGCTGAACCAATTTCTTACAGCAGATTTTCAGCATTTAATCTTGTGATCAAC 4173
Oy 488 aaatatttagcaaatatttatacaagacatccatcagcatttagaattccacgta 547
Db 4174 CAATCAATTTGATGATTCGAATCAAAATCAACCCAGCATGTTTAATGCTTCTACACA 4233
Oy 548 tcacgttacatc 561
Db 4234 GAAATATAGCATC 4247

RESULT 9
X13342/c
ID X13342 standard; DNA; 1479 BP.
AC X13342:
DT 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:405.
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.
PN M09850555-A2.
PD 12-NOV-1998.
PF 04-MAY-1998; U08985.
PR 14-NOV-1997; US-066009.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046555.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PJ, Kunsch CA;
DR WPI: 99-045171/04.
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
PS Claim 1; Page 1617-1618; 2084pp; English.
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC X1938 to X1919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
SQ Sequence 1479 BP; 521 A; 236 C; 316 G; 399 T;

Query Match 3.0%; Score 34.6; DB 1; Length 1479;
Best Local Similarity 44.3%; Pred. No. 2.5;
Matches 136; Conservative 1; Mismatches 170; Indels 0; Gaps 0;

Oy 745 gattcagataacctactcgtgactactcgtttatcagcttcacatgaagaacatg 804
Db 1367 GCTTAGCATTCACAAATTCCTCAATTAACCTTGTGAAATTTCTTCAAAAAGCGCTTT 1308
Oy 805 lactcagatgcacgatacacttggttaacaataatgatatgaacgaatgacacagt 864
Db 1307 TAGTAAATCTGGATGACTACGAGCTTACAAAACCTGTGGGCTTTTGAATTTTTCAAC 1248
Oy 865 tctgttcaaaaataacacacgttgatcagagatcgtgtgacgcgaacattgacatc 924
Db 1247 ACGTATACGTAAACCTTAAAYCCGATCGGTTGCTGTAATAATTCATTGGCATTTGATC 1188
Oy 925 aaatgaagtgaattccgcgaactgagctccacatccatcattacaggaacattctc 984
Db 1167 TTGTTTATAGACTGCTTCGATTTTACATAGTTTACATATAATATACGACGATCTTG 1128
Oy 985 atcataaagaccttcacataaacaatgcttcgaatgacatgattatgaatgat 1044
```

```
Db 1127 ACGTTCACAAATTCCTGCATGATGATTCATTTTCATATTCGTAATATCATATAATAT 1068
Oy 1045 tgttact 1051
Db 1067 TGTTCCT 1061

RESULT 10
V23659/c
ID V23659 standard; cDNA; 2155 BP.
AC V23659:
DT 17-AUG-1998 (first entry)
DE Human interleukin-1 receptor accessory molecule cDNA.
KW Interleukin-1 receptor accessory molecule; IL-1R ACW; human;
KW signal transduction; infection; septic shock; inflammation;
KW rheumatoid arthritis; therapy; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 303..1373
FT sig_peptide 303..353
FT /tag= a
FT /tag= b
FT mat_peptide 354..1370
FT /tag= c
PN M09808969-A1.
PD 05-MAR-1998.
PF 26-AUG-1996; U13954.
PF 26-AUG-1996; WO-013954.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Bednarik DP, Olsen HS, Rosen CA;
DR WPI: 98-230267/20.
DR P-PSDB: W53897.
PT Nucleic acid encoding interleukin-1 receptor accessory protein -
PT used for therapeutic modulation of IL-1 activity
PS Claim 2; Fig 1; 95pp; English.
CC This cDNA clone codes for human interleukin-1 receptor accessory
CC molecule (IL1-R ACW) (see W53897), a new member of the
CC immunoglobulin superfamily that forms a complex with type 1 IL1-R
CC and which has higher affinity for IL-1 than the receptor itself,
CC suggesting that the known high and low affinity forms of IL1-R are
CC in fact the receptor with or without IL1-R ACW, respectively.
CC The 2155 bp sequence is present in clone HME52 (deposited as
CC ATCC 97666) derived from microvascular epithelium (no details of
CC isolation given). Recombinant expression in Escherichia coli,
CC mammalian and insect cells is described. Recombinant host cells
CC and recombinant vectors are claimed. Also claimed are isolated
CC nucleic acid molecules encoding epitope-bearing portions (see
CC W53898-915) of IL1-R ACW. Recombinant IL1-R ACW can be used to
CC identify IL-1R agonists and antagonists useful for therapeutic
CC modulation of IL-1 activity, and to raise specific antibodies.
CC Nucleic acid fragments are useful as diagnostic probes and primers,
CC for isolation of IL1-R ACW-encoding genomic sequences, for in situ
CC hybridisation to determine chromosomal localisation of the gene,
CC for chromosome identification and for identifying mutations that
CC might be linked to disease.
SQ Sequence 2155 BP; 709 A; 429 C; 433 G; 584 T;

Query Match 3.0%; Score 34.6; DB 1; Length 2155;
Best Local Similarity 49.7%; Pred. No. 3;
Matches 88; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

Oy 371 tgcctatttgcttcttaagtatactgcatgcatgcatgataagaagcgtgatcga 430
Db 1974 TTGATACATAGAGTTTAAATAATATATTAATATGCAATTTTGGAAAATAAACATTTAA 1915
Oy 431 tccgcataatttatacgtgagcttccttgccagacatuaaalcgtgagatccaa 490
Db 1914 TACCTTAATTAATTGTCATACATTTGATGAGTGCCTGATTAATAATAGTTTACACATACGA 1855
Oy 491 ttatttagcaaatatttatacaagacatccatcagcatttagaattccacagta 547
Db 1854 TAATGCTTCTAATATATTAATAAGGTATATATTCCTTCTGTATTTGTCACGA 1798
```

```
RESULT 11
ID V74396 standard; DNA: 9062 BP.
AC V74396;
DT 16-MAR-1999 (first entry) SEQ ID #85.
DE Staphylococcus aureus contig SEQ ID #85.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT 721..780
FT misc_feature
FT /*tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature
FT 2521..2580
FT /*tag= b
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature
FT 4321..4380
FT /*tag= c
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature
FT 6121..6180
FT /*tag= d
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature
FT 7921..7980
FT /*tag= e
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
PN EP-786519-A2.
PD 30-JUN-1997.
PR 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
PI WPI: 97-374922/35.
DR Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1: Page 544-549; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
```

```
CC computer readable medium.
SQ Sequence 9062 BP; 2756 A; 1655 C; 1271 G; 3076 T;

Query Match 3.0%; Score 34.6; DB 1; Length 9062;
Best Local Similarity 37.2%; Pred. No. 5.8;
Matches 94; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 819 gatcaacttggtaacaataatgatgaagaatgacacagttgtgtacaaaaa 878
DB 2495 GATCACCCGATTGTGTACAGAGTGCTNNNNNNNNNNNNNNNNNNNNNNNN 2554
QY 879 taccacgtgattcagagatcgtgtacgcgcacatgtcatgaatcaaatggaatgaa 938
DB 2555 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2614
QY 939 ttcgcgcaacttgaagcttcacattcatcttcggcaacctcttgatcatcaaacctt 998
DB 2615 TACCTACACCTTTGCTGTGCTTTTGAATCATGTGCTGATTTGATACACCAACACACA 2674
QY 999 tcacataaacatgctccttcgaaatgcatatgctgtataaatgtatgttatgaagtgt 1058
DB 2675 TTGCATTACAGTGGATTTGCCATCATACGTTTCATCAATTCGATTTCACACGAGTTTG 2734
QY 1059 gtccacatccaat 1071
DB 2735 TTGGAATACCAAT 2747
```

```
RESULT 12
ID X13031 standard; DNA: 21252 BP.
AC X13031;
DT 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:94.
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.
PN W09850555-A2.
PD 12-NOV-1998.
PR 04-MAY-1998; U08985.
PR 14-NOV-1997; US-066009.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046655.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PJ, Kunsch CA;
PI WPI: 99-045171/04.
DR New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
PS Claim 1: Page 615-626; 2084pp; English.
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC X12938 to X13919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosis of Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
SQ Sequence 21252 BP; 6141 A; 4384 C; 3639 G; 7073 T;

Query Match 3.0%; Score 34.6; DB 1; Length 21252;
Best Local Similarity 51.6%; Pred. No. 8.6;
Matches 79; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
```

[illegible]

```

QY      855  tgcacacagttgtgttacaana  877
          | | | | | | |
Db      292  ATCGAGCTGCTGTTTACAAAA  314

```

RESULT	15
T36122/c	
ID	T36122 standard; DNA; 4865 BP.
AC	T36122;
DT	26-OCT-1996 (first entry)
DE	S. mutans antigen I/II gene.
KW	Caries; antigen I/II; epitope; tooth decay; vaccine;
KM	genetic immunisation; ss.
OS	Streptococcus mutans.
FT	Key
FT	location/Qualifiers
FT	100..4785
FT	cds
FT	
FT	/*tag= a
FT	/transl_except= 2001..2049
FT	/note= "Passes 2001-2049 code for an amino acid

FT
TN
PN W09623886-A1.
PD 08- AUG-1996 .
PE 31-JAN-1996; G00207 .
PR 31-JAN-1995; GB-001826.
PA (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PI Kelly C, Lennet T;
DR WPJ: 96-371434/37.
DR P-PADB: W02098 .
PT Nucleic acid encoding polypeptide for prevention of
PT dental caries - which stimulates T or B cell response, and/or
PT adheres to tooth in competition with Streptococcus mutans antigen
PT I/II
PS Disclosure; Page 46-49; 63pp: English.
CC A Streptococcus mutans gene (T36122) codes for I/II antigen
CC (W02098), a 185 kDa cell surface protein at least partly responsible
CC for S. mutans adhesion to teeth. The I/II antigen includes a
CC series of overlapping T-cell, B-cell and adhesion epitopes.
CC Fragments (see also T36111-21) of the gene can be used to produce
CC recombinant polypeptides (W02087-97) carrying such epitopes for
CC use in vaccines for immunisation against dental caries. The DNA
CC fragments can also themselves be used as naked nucleic acid

CC vaccines. 4865 BP; 1687 A; 967 C; 953 G; 1258 T;
SQ Sequence

Query Match 2.9%; Score 34.2; DB 1; Length 4865;
Best Local Similarity. 49.2%; Pred. No. 5.7;
Matches 90; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 379 tggcttcaatgtaatactgactgcatgataaagctgtagatccgcata 438
|||||
DB 1608 TGGTCTGTAGTTCCGCTTCATTTTATGTTCACAGTTCGCCAGTCCAGCTTT 1549
QY 439 ttgtatagctgagcttcttggccagccataatcgtltgatatccaaattattag 498
| | | | |
DB 1548 AATAGACCTGTTCATCTTCGATGCCCTTAACCTGATAGTCTGTAATCTTT 1489
QY 499 caaatattatcaagagccatccatcaagcattagaattccacagatatccgflacc 558
|||||
DB 1488 TTGATATTTGGCAAGTCTGTTGATCTTACGACGTTTGGCTCGTAATCAGCTTTAGC 1429
QY 559 atc 561
|||
DB 1428 ATC 1426

Search completed: April 16, 2000, 04:47:46
Job time: 10165 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 16, 2000, 04:36:46 ; Search time 165.68 Seconds
(without alignments)
838.958 Million cell updates/sec

Title: US-09-323-427-5
Perfect score: 1161
Sequence: 1 tgcctctgattgacgaa.....gcataagacgaatcatcat 1161

Scoring table: IDENTITY_NNC
Gapop 10.0 , Gapext 1.0

Searched: 214294 seqs, 59861574 residues

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2-6/ptodata/1/ina/5A_COMB.seq:*
- 2: /cgn2-6/ptodata/1/ina/5B_COMB.seq:*
- 3: /cgn2-6/ptodata/1/ina/5C_COMB.seq:*
- 4: /cgn2-6/ptodata/1/ina/5D_COMB.seq:*
- 5: /cgn2-6/ptodata/1/ina/6_COMB.seq:*
- 6: /cgn2-6/ptodata/1/ina/PTUS9_COMB.seq:*
- 7: /cgn2-6/ptodata/1/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33.4	2.9	1867	1	US-07-772-087-1
2	33	2.8	11225	7	5182210-9
3	32.6	2.8	5177	7	5352450-1
4	32.4	2.8	576	6	PCT-US96-05320A-705
5	32.4	2.8	4090	4	US-08-781-802-5
6	32	2.8	2338	2	US-08-425-069-1
7	32	2.8	2338	4	US-08-317-844B-1
8	31.6	2.7	4437	2	US-08-559-303B-72
9	31.4	2.7	1189	1	US-07-781-034-4
10	31.4	2.7	1189	6	PCT-US92-08328-4
11	31.4	2.7	3414	2	US-08-764-100-17
12	31.4	2.7	3414	2	US-08-764-100-21
13	31.4	2.7	4970	2	US-08-764-100-14
14	31.4	2.7	4970	2	US-08-764-100-20
15	31.4	2.7	5655	4	US-08-989-478-1
16	31.2	2.7	7295	4	US-08-487-826B-15
17	31.2	2.7	14602	2	US-08-597-236-1
18	31.2	2.7	14602	2	US-08-746-682A-1
19	31	2.7	8855	3	US-08-542-003-1
20	31	2.7	8855	3	US-08-322-760A-1
21	30.8	2.7	325	1	US-08-111-316-6
22	30.8	2.7	325	1	US-08-468-405-6
23	30.8	2.7	326	1	US-08-111-316-7
24	30.8	2.7	326	1	US-08-468-405-7
25	30.8	2.7	1539	3	US-08-828-596-1
26	30.8	2.7	2254	4	US-08-635-066-1
27	30.6	2.6	5049	2	US-08-336-345-1

28	30.6	2.6	5049	2	US-08-336-345-2	Sequence 2, Appli
29	30.6	2.6	5049	3	US-08-647-655-1	Sequence 1, Appli
30	30.6	2.6	5049	3	US-08-647-655-2	Sequence 2, Appli
31	30.6	2.6	19124	4	US-08-487-826B-13	Sequence 13, Appli
32	30.2	2.6	423	1	US-08-470-179-41	Sequence 41, Appli
33	30.2	2.6	2291	2	US-08-920-812-9	Sequence 9, Appli
34	30.2	2.6	2291	2	US-08-920-827-9	Sequence 9, Appli
35	30.2	2.6	2291	2	US-08-921-177-9	Sequence 9, Appli
36	30.2	2.6	2291	2	US-08-362-577C-9	Sequence 9, Appli
37	30.2	2.6	2291	3	US-08-920-828-9	Sequence 9, Appli
38	30	2.6	1478	5	US-08-817-926-1	Sequence 19, Appli
39	30	2.6	3562	5	US-08-817-926-19	Sequence 4, Appli
40	30	2.6	7742	3	US-08-882-704A-4	Sequence 2, Appli
41	29.8	2.6	970	2	US-08-690-095-2	Sequence 2, Appli
42	29.8	2.6	2275	4	US-08-743-637B-2	Sequence 2, Appli
43	29.8	2.6	2275	5	US-08-526-840B-2	Sequence 2, Appli
44	29.8	2.6	2981	2	US-08-257-073-2	Sequence 2, Appli
45	29.8	2.6	2981	3	US-08-184-009-119	Sequence 119, App

ALIGNMENTS

RESULT 1
US-07-772-087-1/c
; Sequence 1, Application US/07772087
; Patent No. 5275945
; GENERAL INFORMATION:
; APPLICANT: HSIAG, Hung-Yu
; APPLICANT: FODGE, Douglas W.
; APPLICANT: LALONDE, James J.
; TITLE OF INVENTION: ALKALINE PROTEASES STABLE IN HEAVY-DUTY
; TITLE OF INVENTION: DETERGENT LIQUIDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/772,087
; FILING DATE: 19911108
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; TELEPHONE/DOCKET NUMBER: 16754/115 CHCO
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1867 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; STRAIN: 164A-1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 716..1849
; OTHER INFORMATION: /product="mature protein of
; OTHER INFORMATION: 164A-1"
; FEATURE:

```

; NAME/KEY: sig_peptide
; LOCATION: 716..1021
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1022..1846
;
US-07-772-087-1

```

Query Match	2.98;	Score 33.4;	DB 1;	Length 1867;
Best Local Similarity	52.58;	Pred. No. 2.2;		
Matches 96;	Conservative 0;	Mismatches 86;	Indels 1;	Gaps 1.

[illegible]

```

RESULT 2
5182210-9
; Patent NO. 5182210
; APPLICANT: BINNS, MATTHEW M.; BOURSNELL, MICHAEL E. G.;
; CAMPBELL, JOAN I. A.; TOMLEY, FIONA M.
; TITLE OF INVENTION: FOMUPOX VIRUS PROMOTERS
; NUMBER OF SPOUNCES: 22
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,608
; FILING DATE: 21-OCT-1988
; SEQ ID NO.: 9
; LENGTH: 11225
5182210-9

```

Query Match	2.88;	Score 33;	DB 7;	Length 11225;
Best Local Similarity	55.68;	Pred. No. 7.3;		
Matches	85;	Conservative	0;	Mismatches 65;
				Indels 3;
				Gaps 1

QY	720	caatttcagttgaaagcaggttgatattccatattcccaatcgtgacaaactgtt	779
Db	9023	caatttcagttctatagcagcatcatctatctcgttaacgttatcttcgttatata	9082
QY	780	tatcaggttcctatgtaaaagcattgacacgaatgacagatcaactttggtgaacaaa	839
Db	9083	tatctcgttccatagaatattgtctctttct---aatacctcttcagcttcttgtaaat	9139
QY	840	atggatgaaacgaatgacacacagttggtgttta	872
Db	9140	aagttactcaatttactcgtataatagaatgtttca	9172

RESULT 3
5352450-1/c
Patent No. 5352450
APPLICANT: KOGA, TOSHIHIKO, OKANASHI, NOBUO, TAKANASHI, ICHIRO
SHIBUYA, KOJI; OHTA, HIROTKA
TITLE OF INVENTION: METHOD FOR PREPARING VACCINE FOR DENTAL
CARIES AND VACCINAL COMPOSITIONS FOR DENTAL CARIES USED AS NASAL
DROP
NUMBER OF SEQUENCES: 2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/529,602
FILING DATE: 29-MAY-1990

```

;SEQ ID NO:1
;      LENGTH: 5177
5352450-1

```

Query Match	2.88;	Score 32.6;	DB 7;	Length 5177;
Best Local Similarity	48.68;	Pred. No. 6.4;		
Matches 89;	Conservative 0;	Mismatches 94;	Indels 0;	Gaps 0;

QY	379	tggtctcttcaatgaatgatactatgctcggagatcgatgaagaacgtctggatcgcgtccgata	438
Db	1708	TGGTCTCTGTTAAGTTCCGCTTCATTTTTATGTTTTTCAGTTCGCCAGGACAGCTTT	1649
QY	439	cttgatacgtgagctctcttggccagccatlaatcctgttgatattcccaatlaataag	498
Db	1648	AATGGAAGTTTGTCATCTTCGATGCCCTTTAACTTAACGTGATGTCTGGCTAAATCTTT	1589
QY	499	caaatattatcaagagcacatccatccagcatttgaatttccacagatatacgcgttac	558
Db	1588	TTCGATTATTTGGCAGAGATCGCTGATATCTTAGCAAACTTTTGGTTGTGTAATCAGCTTTAGC	1529
QY	559	atc 561	
Db	1528	ATC 1526	

RESULT 4
PCT-US96-05320A-705
: Sequence Information PC/TUS9605320A
: Sequence Information PC/TUS9605320A
: Sequence Information PC/TUS9605320A

? APPLICANT: Human Genome Sciences
 ? APPLICANT: 9410 Key West Avenue
 ? APPLICANT: Rockville, MD 20850
 ? APPLICANT: United States of America
 ? APPLICANT: Johns Hopkins University
 ? APPLICANT: 720 Rutland Avenue
 ? APPLICANT: Baltimore, MD 21205
 ? APPLICANT: United States of America
 ? APPLICANT: Mark D. Adams
 ? APPLICANT: Owen White
 ? APPLICANT: Hamilton O. Smith
 ? APPLICANT: J. Craig Venter
 ? TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome
 ? NUMBER OF SEQUENCES: 48
 ? CORRESPONDENCE ADDRESS:

CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20003-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/05320A
FILING DATE: April22, 1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: June 7, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: June 7, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Eric K. Steffe
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.014PC01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 571-2600
TELEFAX: (202) 371-2540

```

; INFORMATION FOR SEQ ID NO: 705:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 576 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;
; PCT-US96-05320A-705
;
Query Match          2.8%; Score 32.4; DB 6; Length 576;
Best Local Similarity 56.6%; Pred. No. 2.4;
Matches 60; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
;
Oy 19 aaattaaacgaatgtcaataatgaacgacgaacatcaatgaatgtcaacccaataa 78
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 81 AAGTTTATCGAGGCGATATTTTTCCTATTCGCCGCGATGATGATGATGCCAATGTC 140
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Oy 79 catgagaagccaatgtgtacatgcagattccatttctactgca 124
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 141 TATGTCAAGCAAAAGTGGCGTGAAATTTCATTATATACGCCA 186
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 5
US-08-781-802-5
; Sequence 5, Application US/08781802
; Patent No. 5969121
; GENERAL INFORMATION:
;   APPLICANT: ALLEN, Larry
;   APPLICANT: AIKENS, John
;   APPLICANT: FONSTEIN, Michael
;   APPLICANT: VONSTEIN, Veronika
;   APPLICANT: DEMIRJIAN, David
;   APPLICANT: CASADABAN, Malcolm
;   TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
;   NUMBER OF SEQUENCES: 12
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
;     STREET: 300 S. Wacker Drive 32nd Floor
;     CITY: Chicago
;     STATE: Illinois
;     COUNTRY: USA
;     ZIP: 60606
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: Patentln Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/781.802
;     FILING DATE: 10-JAN-1997
;     CLASSIFICATION: 536
;     TITLE OF INVENTION:
;     NUMBER OF SEQUENCES: 12
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 08/694,078
;       FILING DATE: 07-AUG-1996
;       CLASSIFICATION: 536
;       PRIOR APPLICATION DATA:
;         APPLICATION NUMBER: US 60/019,580
;         FILING DATE: 12-JUN-1996
;         PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: US 60/009,704
;           FILING DATE: 11-JAN-1996
;           PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US 60/001,995
;             FILING DATE: 01-AUG-1995
;             ATTORNEY/AGENT INFORMATION:
;               NAME: Chao, Mark
;               REGISTRATION NUMBER: 37,293
;               REFERENCE/DOCKET NUMBER: 95,963-E
;               TELECOMMUNICATION INFORMATION:
;                 TELEPHONE: 312-913-0001
;                 TELEFAX: 312-913-0002
;             INFORMATION FOR SEQ ID NO: 5:
;             SEQUENCE CHARACTERISTICS:

```

```

;   LENGTH: 4090 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 197..1699
;     OTHER INFORMATION: /note="E011 sequence of longest
;     OTHER INFORMATION: open reading frame; other possible start codons ATG/met4;
;     OTHER INFORMATION: TTG/Leu; GTG/Val8; GTG/Val15; GTG/Val36; ATG/met62"
;   FEATURE:
;     NAME/KEY: mat_peptide
;     LOCATION: 197..1699
;
US-08-781-802-5
;
Query Match          2.8%; Score 32.4; DB 4; Length 4090;
Best Local Similarity 44.5%; Pred. No. 6.5;
Matches 129; Conservative 0; Mismatches 161; Indels 0; Gaps 0;
;
Oy 723 tttaggttgaagaagcagttgtgattcagataacctaatctgtgcactaactgtttat 782
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 2606 TCTGATTAAACAAGATATTGTGGCGCGAAACAGCCGGCGTTCTCTCCCTCCCTCTT 2665
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Oy 783 cagctccatgtaaagcagctgtactcgatagacagcagcaacttggtaacaataatg 842
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 2666 GATCCACTCTATTATTGCCCCCTACAGGCTTACAAAACAATTCTTTTGTAACTATATAA 2725
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Oy 843 gatgaacgaatgaacacagctgtgtacaaaataacacagctgattcagagatcgtg 902
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 2726 GATAAATGCGCAACCCCTTATTACAGAGGCTTGGCGATCGGATATTCCACACA 2785
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Oy 903 taagcgaacattgtcatcaatcaatgaatgaatgaatcgcgcaactgtacgtccact 962
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 2786 TTTTTCATCTGTCTGAGAAGTCAATGCTGCTCACCACAGTTAACTGTTAATGAGATA 2845
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Oy 963 catcatcgcgaacctcttgatcataaagaccttcacataaagatgt 1012
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 2846 TCAATCTGCTGCTATTCTTCTCACCATATGAGCGTTATTCAAAAACAATGT 2895
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 6
US-08-425-069-1/c
; Sequence 1, Application US/08425069
; Patent No. 5728810
; GENERAL INFORMATION:
;   APPLICANT: Lewis, Randolph V.
;   APPLICANT: Xu, Ming
;   APPLICANT: Himan, Michael B.
;   TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
;   TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
;   CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
;   NUMBER OF SEQUENCES: 69
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Birch, Stewart, Kolasch & Birch
;     STREET: 301 No. 5728810th Washington Street
;     CITY: Falls Church
;     STATE: Virginia
;     COUNTRY: U.S.A.
;     ZIP: 22046
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: Patentln Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/425,069
;     FILING DATE: 19-APR-1995
;     CLASSIFICATION: 435
;     ATTORNEY/AGENT INFORMATION:
;       NAME: Murphy Jr., Gerald M
;       REGISTRATION NUMBER: 28,977

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 16, 2000, 03:09:15 ; Search time 3225.5 Seconds
(without alignments)
1359.028 Million cell updates/sec

Title: US-09-323-427-5
Perfect score: 1161
Sequence: 1 tgcctctgattcgacgaa.....gcataagacgacatcatcat 1161

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*

- 1: em_est1:*
- 2: em_est2:*
- 3: em_est3:*
- 4: em_est4:*
- 5: em_est5:*
- 6: em_est6:*
- 7: em_est7:*
- 8: em_est8:*
- 9: em_est9:*
- 10: em_est10:*
- 11: em_est11:*
- 12: em_est12:*
- 13: em_est13:*
- 14: em_est14:*
- 15: em_est15:*
- 16: em_est16:*
- 17: em_est17:*
- 18: em_est18:*
- 19: em_est19:*
- 20: gb_est1:*
- 21: gb_est2:*
- 22: gb_est3:*
- 23: gb_est4:*
- 24: gb_est5:*
- 25: gb_est6:*
- 26: gb_est7:*
- 27: gb_est8:*
- 28: gb_est9:*
- 29: gb_est10:*
- 30: gb_est11:*
- 31: gb_est12:*
- 32: gb_est13:*
- 33: gb_est14:*
- 34: gb_est15:*
- 35: gb_est16:*
- 36: gb_est17:*
- 37: gb_est18:*
- 38: gb_est19:*
- 39: gb_est20:*
- 40: gb_est21:*
- 41: gb_est22:*
- 42: gb_est23:*
- 43: gb_est24:*
- 44: gb_est25:*

45: gb_est26:*

46: gb_est27:*

47: gb_est28:*

48: gb_est29:*

49: gb_est30:*

50: gb_est31:*

51: gb_est32:*

52: gb_est33:*

53: gb_est34:*

54: gb_est35:*

55: gb_est36:*

56: gb_est37:*

57: gb_est38:*

58: gb_est39:*

59: gb_est40:*

60: gb_est41:*

61: gb_est42:*

62: gb_est43:*

63: gb_est44:*

64: gb_est45:*

65: gb_est46:*

66: gb_est47:*

67: gb_est48:*

68: gb_est49:*

69: gb_est50:*

70: gb_est51:*

71: gb_est52:*

72: gb_est53:*

73: gb_est54:*

74: gb_est55:*

75: gb_est56:*

76: gb_est57:*

77: gb_est58:*

78: gb_est59:*

79: gb_est60:*

80: gb_est61:*

81: gb_est62:*

82: gb_est63:*

83: gb_est64:*

84: gb_est65:*

85: gb_est66:*

86: gb_est67:*

87: gb_est68:*

88: gb_est69:*

89: gb_est70:*

90: gb_est71:*

91: gb_est72:*

92: gb_est73:*

93: gb_est74:*

94: gb_est75:*

95: gb_est76:*

96: gb_est77:*

97: gb_est78:*

98: gb_est79:*

99: gb_est80:*

100: gb_est81:*

101: gb_est82:*

102: gb_est83:*

103: gb_est84:*

104: gb_est85:*

105: gb_est86:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	390.2	33.6	665	42	AI111196 SMOV3MCA1
C 2	372.2	32.1	628	37	AA701731 SMOV3MCA1

C	3	364.2	31.4	672	36	AA618895	AA618895 SMOV3MCA1
C	4	358.6	30.9	639	44	AI322117	AI322117 SMOV3MCA1
C	5	355.8	30.6	610	37	AA668071	AA668071 SMOV3MCA1
C	6	324.2	27.9	776	35	AA585626	AA585626 SMOV3MCA1
C	7	312.4	26.9	537	37	AA668051	AA668051 SMOV3MCA1
C	8	306.8	26.4	552	44	AI322068	AI322068 SMOV3MCA1
C	9	299.6	25.8	615	44	AI317885	AI317885 SMOV3MCA1
C	10	291.2	25.1	531	39	AA901444	AA901444 SMOV3MCA1
C	11	278.8	24.0	363	36	AA625020	AA625020 SMOV3MCA1
C	12	269.2	23.2	354	39	AA841200	AA841200 SMOV3MCA1
C	13	267.8	23.1	466	36	AA625024	AA625024 SMOV3MCA1
C	14	263	22.7	459	36	AA618952	AA618952 SMOV3MCA1
C	15	260.4	22.4	395	36	AA625010	AA625010 SMOV3MCA1
C	16	257	22.1	369	36	AA625022	AA625022 SMOV3MCA1
C	17	246.8	21.3	528	44	AI322078	AI322078 SMOV3MCA1
C	18	241.2	20.8	375	36	C71078	C71078 SMOV3MCA1
C	19	237.2	20.4	379	62	AV203892	AV203892 SMOV3MCA1
C	20	235.8	20.3	360	36	C67783	C67783 SMOV3MCA1
C	21	233.4	20.0	377	36	C70345	C70345 SMOV3MCA1
C	22	222.2	19.1	493	74	AA208357	AA208357 SMOV3MCA1
C	23	215	18.5	504	36	AA624955	AA624955 SMOV3MCA1
C	24	214	18.4	504	36	AA624955	AA624955 SMOV3MCA1
C	25	210.8	18.2	360	62	AV196275	AV196275 SMOV3MCA1
C	26	210.2	18.1	377	62	AV186785	AV186785 SMOV3MCA1
C	27	207.2	17.8	360	62	AV202546	AV202546 SMOV3MCA1
C	28	205.6	17.7	322	36	AA625021	AA625021 SMOV3MCA1
C	29	205.4	17.6	360	62	AV201206	AV201206 SMOV3MCA1
C	30	204.8	17.6	360	62	AV191120	AV191120 SMOV3MCA1
C	31	184.2	15.9	417	36	AA625044	AA625044 SMOV3MCA1
C	32	173.2	14.9	443	36	AA625040	AA625040 SMOV3MCA1
C	33	163.2	14.1	281	46	AI438781	AI438781 SMOV3MCA1
C	34	162.2	14.0	318	20	T01970	T01970 SMOV3MCA1
C	35	159	13.7	380	44	AI322108	AI322108 SMOV3MCA1
C	36	146.2	12.6	635	44	AI317939	AI317939 SMOV3MCA1
C	37	144	12.4	522	44	AI322073	AI322073 SMOV3MCA1
C	38	140.2	12.1	317	45	AI352993	AI352993 SMOV3MCA1
C	39	138.2	11.9	467	31	AA294218	AA294218 SMOV3MCA1
C	40	119	10.2	273	49	AI655774	AI655774 SMOV3MCA1
C	41	103.4	8.9	473	36	AA625005	AA625005 SMOV3MCA1
C	42	102.6	8.8	499	44	AI313768	AI313768 SMOV3MCA1
C	43	95.4	8.2	470	44	AI313804	AI313804 SMOV3MCA1
C	44	87.6	7.5	409	36	AA625029	AA625029 SMOV3MCA1
C	45	81	7.0	407	36	AA625007	AA625007 SMOV3MCA1

ALIGNMENTS

RESULT 1
LOCUS A111196 665 bp mRNA EST 31-AUG-1998
DEFINITION SMOV3MCA1232SK Onchocerca volutus molting L3 larva cDNA
(SL96MLM-Ovml3) Onchocerca volutus cDNA clone SMOV3MCA1232 5',
mRNA sequence.

ACCESSION A111196
VERSION A111196.1 GI:3510080
KEYWORDS EST.
SOURCE Onchocerca volutus.
ORGANISM Onchocerca volutus
Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;
Filarioidea; Onchocercidae; Onchocerca.

REFERENCE 1 (bases 1 to 665)
AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
TITLE Genes expressed in molting L3 larvae of Onchocerca volutus
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1407460.
Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786

Email: genome@smith.edu

Seq primer: pBluescript SK.

Location/Qualifiers

1..665

/organism="Onchocerca volutus"

/strain="Kumba, Camerons"

/db_xref="taxon:6282"

/clone="SMOV3MCA1232"

/clone_lib="Onchocerca volutus molting L3 larva cDNA (SL96MLM-Ovml3)"

/dev_stage="molting L3"

/lab_host="XLI-Blue MR"

/note="Vector: lambda Uni-Zap XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (forest strain). The L3 were cultured in 20% FCS in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volutus molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda Uni-Zap XR vector and has 1 x 10E6 independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustigman@bc.org)."

BASE COUNT

186 a 129 c 136 g 213 t 1 others

ORIGIN

Query Match

Best Local Similarity

Matches 451; Conservative

33.6%; 84.1%; 0;

Score 390.2; DB 42; Length 665; Pred. No. 1.3e-94; Mismatches 84; Indels 1; Gaps 1;

Oy

577

atgagcaacgcgcgagaaagatcaacggttcagaatgcagatgcatgataaac

636

Db

651

atgtccaccagcccgcaatgattcattcattcattcattcattcattcattc

593

Oy

637

tggcgcgacaaatgataagaaatgacgtgtgacgggttgccacatcaaatlc

696

Db

592

tggttgaccagatgagaaatgacaggtgtccgacggtccacacgaatctc

533

Oy

697

ataacgcatctgcatcgacgcgacacattgagttgcaaaagagtgatcagatc

756

Db

532

atattccgcatfaccagcattccgacacatttgcattcattcattcattcattc

473

Oy

757

ctcaatctgacactaactgtttatcagcttcacatgtaaaagcatgtaacatgc

816

Db

472

tttcagattgagcgcgcttactgttcttccatgcttccatgcttccatgctt

413

Oy

817

acgatacacttctgtaacaaataatgataagaaatgacaaagatgtgtgtacaa

876

Db

412

tcgattcaactttttgacaaacaaatgattgattgattgattgattgattgatt

353

Oy

877

aatacaccgtgattcagagatcgtgtacgcgcaacattgcatgataaataagtg

936

Db

352

aataccacgctgattttaaaccagacgctgacgcttaccatgacgacgacgac

293

Oy

937

aattccgcaacttgcagtcacacattcaacacacacacacacacacacacacac

996

Db

292

aattccgctgacgctgacgctgacgctgacgctgacgctgacgctgacgctgac

233

Oy

997

tttcacataaagatgcttctgtaagatgcttctgtaagatgcttctgtaagatg

1056

Db

232

ttttacgattatgattccttcaaaatgattgattgattgattgattgattgattg

173

Oy

1057

tgttcacatcaaatcttgcctacacacacacacacacacacacacacacacac

1112

Db

172

tggaccacattcatttctggttacccttcttaccacgcttaccacgacgacgac

117

RESULT	2
AA701731/c	
LOCUS	
DEFINITION	AA701731 628 bp mRNA EST 19-DEC-1997 SMOV3MCA1144SK Onchocerca volvulus molting L3 larva cDNA (SL96MLW-OVWL3) Onchocerca volvulus cDNA clone SMOV3MCA1144 5'
ACCESSION	AA701731
VERSION	AA701731.1 GI:2704931
KEYWORDS	EST.
SOURCE	Onchocerca volvulus.
ORGANISM	Onchocerca volvulus. Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spirurida Filarioidea; Onchocercidae; Onchocerca. 1 (bases 1 to 628) Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S. Genes expressed in molting L3 larvae of Onchocerca volvulus Unpublished (1997)
JOURNAL	On Sep 19, 1997 this sequence version replaced gi:1517220.
COMMENT	Contact: Steven A. Williams Molecular Parasitology Smith College Department of Biological Sciences Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA Tel: 4135853826 Fax: 4135853786 Email: genome@smith.edu Seq primer: pblunescript SK. Location/Qualifiers
FEATURES	

```

source
1. .628
/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMOV3MCA114"
/clone_lib="Onchocerca volvulus molting L3 larva cDNA
(SL96MLM-OvML3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/note="Vector: lambda Uni-ZAP XR; Site 1: Eco RI; Site 2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ML3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNA pol I. The library was constructed in the lambda
Uni-ZAP XR vector and has 1 x 10E6 independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Maniewski in the laboratory of Dr. S. A. Williams
The library is available from Dr. Sara Lustigman (email:
slustigman@nbc.org)."
181 a 121 c 125 g 201 t

```

	Query Match	32.1%	Score 372.2	DB 37	Length 628	
	Best Local Similarity	84.3%	Pred. No. 9.3e-90			
	Matches 431	Conservative 0	Mismatches 78	Indels 2	Gaps 1	
Oy	604 gtttcagaatcgcgatgcatatttgata--aaactgtctgacccaatgatgaacaattga	661 				
Db	622 GCTTCAACAATCACACATGTCATTATGTGTAACTGGGTTGAACATAGTACAAATTCA	563 				
Oy	662 actgtgttacacggtttgcgtccaccatcccaaatlttaaacgcatcatctgcggagaca	721 				
Db	562 ACAGGTTCTCCGGATGGTCCACCATCAAGAATTTATATACCGCATACAGGATCGGTACA	503 				
Oy	722 atttagagttaaaggcggttgttgatttcagttacctcaactctgtgcactaactgttta	781 				

Db	502	ATTGTGTGCAAAATGACAGTGTGTGATTTTCGGAAATTCACGTTGAGCGCTTAACGTGTCTTA	443
Qy	782	tcagcttccatcgtlaaagaacgatctgtacttcgatatacgacgatacacttctgtgtaacaataat	841
Db	442	TCAGCTTTCACATGTAAACACATTTGTATTCGAATATGCTCGATCAACTTTTGTCCACAACACAT	383
Qy	842	ggatgaaaccgaaatgacacacgttctgttctacaaaataaccacgttgatctcaagatcgt	901
Db	382	GGATGCAATGTGAGTACACAAACATGATGTGACAAAATAACACAGCGGATTTTAACCAACAGT	323
Qy	902	gtacgcgcaacatctgcataatcaaatggaagtgaatttcgcgcaacttgaacgtccacact	961
Db	322	GTACGTGCTTAATTTACAGAGAAATCAACCGAAGTTCAATTCGGCTACTGTGACGTCCACCT	263
Qy	962	tcatcatcagcgcaacacctctctgtatcacaataaagaaccttcacataaacaatgtccttcgat	1021
Db	262	TCATCATTTTCGGCATTCCTGCGCTGATTCGTAAGAAGCTTTACAGGTATACATGTGTCCTTCAAAAT	203
Qy	1022	gcataatcgtatataaataatctgaatctgtatctgaagtctgcacatctcaatcttcgtgtca	1081
Db	202	GGATTTACAGCATTTAAAGCTGTGACACGTATTTGATGTGGACACACATTCATTTCTGTGTCA	143
Qy	1082	ccttcgacacacatctgtcacaacggaatcgaat	1112
Db	142	CCTTCTACACCGTTATTCACCGGATTAACAT	112

RESULT 3
 AA618895/c
 LOCUS
 DEFINITION
 AA618895 672 bp mRNA EST 12-NOV-1997
 SMOV3MCAT1879SK Onchocerca volvulus molting L3 larva cDNA
 (S1596LW-Ovmlu3) Onchocerca volvulus CDNA clone SML3CO1879 5', mRNA
 sequence.
 ACCESSION
 AA618895
 VERSION
 AA618895.1 GI:2522771
 KEYWORDS
 EST.
 SOURCE
 Onchocerca volvulus.
 ORGANISM
 Onchocerca volvulus
 Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida;
 Filarioidea; Onchocercidae; Onchocerca.
 1 (bases 1 to 672)
 REFERENCE
 Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
 Authors
 Genes expressed in molting L3 larvae of Onchocerca volvulus
 TITLE
 Unpublished (1997)
 JOURNAL
 On Sep 12, 1996 this sequence version replaced gi:1405188.
 COMMENT
 Contact: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genome@smith.edu
 FEATURES
 Seq primer: pbluescript SK.
 Location/Qualifiers
 1..672

```

//clone="SMML3C01879"
//clone_1id="Onchocerca volvulus molting L3 larva cDNA
(SL96MLM-Ovml3)"
//dev_stage="molting L3"
//lab_host="XLI-Blue MRF"
//note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using

```

reverse transcriptase and Oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda Uni-Zap XR vector and has 1 x 10⁶ independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustigman@bc.org)."

BASE COUNT 198 a 132 c 129 g 202 t 11 others
ORIGIN

Query Match 31.4%; Score 364.2; DB 36; Length 672;
Best Local Similarity 81.1%; Pred. No. 1.3e-87;
Matches 417; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

Qy 599 tcaacggttaagatgcacatgcatcattatgataaactgctgcacaaatgataaact 658
Db 672 tcaactgtttcagaaatcannatntatgaaacactggtngacaaatgattacaaat 613
Qy 659 tgaactgttgacggtgtgaccacacacaaatcacaacgcatcgtgcacg 718
Db 612 GAACGAGTGTCCGGGAGTGCACCATCAAGAATCTCATTCGGCATACAGCATCGGT 553
Qy 719 acaattgagttgaagaacggtgtgattcgaatacctcaatctgtgcactaactgt 778
Db 552 ACAATTGTGTGNGCAAAAGCAGTTGTCTCGAAATTTCAAGTTGAGCCCTAATCTGC 493
Qy 779 ttatcagttcattgataaagcattgtactgataatgcacacacttctgttaacaa 838
Db 492 ttatcagttcattgataaagcattgtactgataatgcacacacttctgttaacaa 433
Qy 839 aatgagatgaagaatgataacagttgttctataaataaccacgtgtgattcagat 898
Db 432 AATGAGTGAATGAGNNCAACAACATGATGTGACAAAATAACACGATGATTAAACGA 373
Qy 899 cgtgtaagcgcaacatcgtcagatcaatgaatgaatgaatcggcaactgtgcac 958
Db 372 CGTGTACTGCTACATTTACAGAAATCAACGAACTTCCGCTACCTGACGTCA 313
Qy 959 ccttcacatcagcgaacactctctgacataaagacacttcacataacatgtcttcg 1018
Db 312 CCTTCATATTTCCGCAATCCCTGATCGTATAGCCTTTCACGTATACATGCTTCA 253
Qy 1019 aatgacatcgtgataaataatgattgtatgaagttgtgcacatcaattctggc 1078
Db 252 AATGAGTGAATGAGTGAATGAGTGAATGAGTGAATGAGTGAATGAGTGAATGAGT 193
Qy 1079 tcaacttcgacacacatcgttcacacggaatgaat 1112
Db 192 TCACCTCTACACCGTTATCAACCGAATAGCAT 159

RESULT 4
AI322117/c 639 bp mRNA EST 22-DEC-1998
LOCUS SMOV3MCA12608K Onchocerca volvulus molting L3 larva cDNA
DEFINITION (SL96MLM-Ovml3) Onchocerca volvulus CDNA clone SMOV3MCA12608 5',
mRNA sequence.
ACCESSION AI322117
VERSION AI322117.1 GI:4056268
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus
Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spiruridae;
Filarioidea; Onchocercidae; Onchocerca.
REFERENCE 1 (bases 1 to 639)
AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL Unpublished (1997)
COMMENT On Jan 14, 1998 this sequence version replaced gi:11797212.
Contact: Steven A. Williams
Molecular Parasitology

Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK.
Location/Qualifiers
1. 639
/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMOV3MCA12608"
/clone_1db="Onchocerca volvulus molting L3 larva cDNA
(SL96MLM-Ovml3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/note="Vector: lambda Uni-Zap XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (mu3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and Oligo(dT) followed by RNase H
and DNA pol I. The library was constructed in the lambda
Uni-Zap XR vector and has 1 x 10⁶ independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustigman@bc.org)."

FEATURES
source

BASE COUNT 197 a 130 c 121 g 191 t
ORIGIN

Query Match 30.9%; Score 358.6; DB 44; Length 639;
Best Local Similarity 84.5%; Pred. No. 4.2e-86;
Matches 403; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 636 ctgctgcacacatgataagcaattgaaatggttgacggtgtgacacacacaaatt 695
Db 638 CTGCTGACCAATGGTATCAATGTAACAGGTTGTCGGAGTGCACCAATCAAGAACT 579
Qy 696 cataacgacacacggtgcacacacacacacacacacacacacacacacacacacac 755
Db 578 CATATCGGATACAGGATCGGTACAAATTTGTGTTGCAAAAGCAGTTGTCTTCGAAA 519
Qy 756 cctcaatctgtgac 815
Db 518 CTTCAAGTTGAGCGCTACCTGCTTATCAGCTTCATGTAACCAATGATGATGATGAT 459
Qy 816 cagatcaacttggtaacaataatgataatgaagaacacacacacacacacacacac 875
Db 458 CTCGATCAACTTTGTGTCACAAACATGATGATGATGATGATGATGATGATGATGAT 399
Qy 876 aaatacacacggtgac 935
Db 398 AAATACGACGTTGATTTACGACGATGATGATGATGATGATGATGATGATGATGAT 339
Qy 936 aaatcgcgcaacttgacgtccacacacacacacacacacacacacacacacacacac 995
Db 338 CAATTCGCGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 279
Qy 278 ctccacataac 1055
Db 219 CTTTCACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 219
Qy 1056 ttgttcac 1112
Db 218 TTGACACATTCATTTCTGTTACCTTCTACACCGTTATCAACCGAATAGCAT 162

RESULT 5
AA668071/c 610 bp mRNA EST 20-NOV-1997
LOCUS
DEFINITION SMOV3MCAM02C01SK Onchocerca volvulus molting L3 larva cDNA (SL96MLW-Ovml3) Onchocerca volvulus cDNA clone SMOV3MCAM02C01 5', mRNA sequence.
ACCESSION AA668071
VERSION AA668071.1 GI:2629570
KEYWORDS Est.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida; 1 (bases 1 to 610)
AUTHORS Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL Unpublished (1997)
COMMENT On Oct 30, 1997 this sequence version replaced gi:2160807.
CONTACT Steven A. Williams
MOLECULAR PARASITOLOGY Smith College Department of Biological Sciences
DEPARTMENT Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA
TEL Tel: 4135853826
FAX Fax: 4135853786
EMAIL Email: genome@smith.edu
SEQ PRIMER Seq primer: Bluescript SK.
FEATURES Location/Qualifiers
source 1..610
/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMOV3MCAM02C01"
/clone_1lb="Onchocerca volvulus molting L3 larva cDNA (SL96MLW-Ovml3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRP"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (forest strain). The L3 were cultured in 20% FCS in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volvulus molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by Rnase H and DNA pol I. The library was constructed in the lambda Uni-Zap XR vector and has 1 x 10E6 independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustig@nbc.org)."
BASE COUNT 187 a 106 c 138 g 178 t 1 others
ORIGIN
Query Match 30.6%; Score 355.8; DB 37; Length 610;
Best Local Similarity 74.0%; Pred. No. 2.4e-85;
Matches 450; Conservative 0; Mismatches 158; Indels 0; Gaps 0.
395 ataccgacatgcgtatgatagaagaagcgtgtgatcgatccgcgatattgtatcgttagct 454
||||| ||||||| |||| ||| ||| ||||| ||||| ||||| ||||| ||||| |||||
609 ATACTCAATCTGGCATTGCGAATTAATTAATCTCAAGATCCATCGCTATTATTAAACATGAGCT 550
445 tctttgccagccatataactctgttgatattccaaattatttgacaattattatcaaga 514
||||| ||||||| ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| |||||
549 TCTTGGCCAGCAGCATTAAGATTGATGATATTCCTCAATTTGTTAGTAAATACCTTGTCACA 490
515 gcaatcctacatcgcatatttgaaattccacagtatataccggttaccatcatcagcaaaagcag 574

DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	FEATURES																					
Db	489	GCACACACCTTCTTCATTCAGAAATCTCTACCGTATACCTTTCCGTATCCACAGTACAA	430																															
Qy	575	gaatggaacaccgcgcgcagaagaatlatcaacggtttcaagaatcgcagtctcatltagataa	634																															
Db	429	GAATGCACATATGACACAGAAATATATCTGTAGTTCGTGAATCACAAGTCATTTGTGTATAT	370																															
Qy	635	actgctcgtaccaaataatagcaaatgtgaactggtttgacgggttgcgtccacatccaataat	694																															
Db	369	ACGTGATATCCCATCATTCATATGCAATGCAATGCGTATACAGATGCGTGCACACTTCGCAATTC	310																															
Qy	695	tcatlaagcgcataccgcgcgtcgcgcgaacatttgagtcttgaagaagcaggttgltgattcaat	754																															
Db	309	TGCTATTCGACACACAGGCAATTTGTAACACTTGTGTCTGGAATGCTGTGCTATTCCTGGAT	250																															
Qy	755	acctcaatctgtcgcctcaactcgttttlatcaagcttccatctgataaagaacatgctactgcgat	814																															
Db	249	ACTTCAGATTTCTGCACATAACGGSTTTTGTTCAGCTTCATATAGAAACACATGTAACGATAT	190																															
Qy	815	gcagcagataactcttgtaacaataatgataatgataaagcaagaatgataacagctgttcttaca	874																															
Db	189	GCTCGCTCATTTTCTTTTGTATATGAAATTCGGTAAATGAAATACAAATTAATGCTGTATACA	130																															
Qy	875	aaataaccacgctggaattcgaagatcgtgtacgcgcgaacatctgcatgaatccaatggaagt	934																															
Db	129	AAGACACCTTCGAGATTTTAGAGACCGTGATCGTTTAACTTAACATTAACATGAATCATTCTGAAC	70																															
Qy	935	gaatctcgcgcgaacttgagctccacacctatcatatcgaagcgaacactcttgcatacataaaga	994																															
Db	69	TCGATTCACGATCTACGCGCTCCACACATCTGTGAACGACATTCATTCGGATCGTAAAGC	10																															
Qy	995	ccttctac 1002																																
Db	9	CCTTTTAC 2																																
RESULT 6																																		
AA585626/c																																		
LOCUS	AA585626	776 bp	mRNA	EST	30-DEC-1997																													
DEFINITION	SW3D9CA349SK Brugia malayi L3 molting-day 9 larva cDNA (SAW97MLM-Bml3d9) Brugia malayi cDNA clone SW3D9CA349 5', mRNA sequence.																																	
ACCESSION	AA585626																																	
VERSION	AA585626.1 GI:2393038																																	
KEYWORDS	EST.																																	
SOURCE	Brugia malayi.																																	
ORGANISM	Brugia malayi.																																	
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Filarioidea; Onchocercidae; Brugia.																																	
AUTHORS	1 (bases 1 to 776)																																	
TITLE	Williams,S.A. and Lizotte-Waniewski,M.																																	
JOURNAL	Genes expressed in day 9 post-infection, third stage larvae of Brugia malayi																																	
COMMENT	Unpublished (1997)																																	
On May 5, 1995 this sequence version replaced gi:797827.																																		
Contact: Steven A. Williams																																		
Molecular Parasitology																																		

AI322066/c 552 bp mRNA EST 22-DEC-1998
 LOCUS SMOV3MCAM12A04SK Onchocerca volvulus molting L3 larva cDNA
 DEFINITION (SL96MLM-Ovml3) Onchocerca volvulus cDNA clone SMOV3MCAM12A04 5', mRNA sequence.
 ACCESSION AI322068
 VERSION AI322068.1 GI:4056219
 KEYWORDS EST.
 SOURCE Onchocerca volvulus.
 ORGANISM Onchocerca volvulus.
 Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida; Filarioidea; Onchocercidae; Onchocerca.
 REFERENCE 1 (bases 1 to 552)
 AUTHORS Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
 TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
 JOURNAL Unpublished (1997)
 COMMENT On Jan 14, 1998 this sequence version replaced gi:11797124.
 CONTACT: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genome@smith.edu
 Seq primer: pBluescript SK.
 FEATURES
 source
 1..552
 /organism="Onchocerca volvulus"
 /strain="Kumba, Cameroons"
 /db_xref="taxon:6282"
 /clone="SMOV3MCAM12A04"
 /clone_1lb="Onchocerca volvulus molting L3 larva cDNA (SL96MLM-Ovml3)"
 /dev_stage="molting L3"
 /lab_host="XLI-Blue MRF"
 /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (forest strain). The L3 were cultured in 20% FCS in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volvulus molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda Uni-Zap XR vector and has 1 x 10E6 independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustigm@mc.org)."
 BASE COUNT 154 a 106 c 111 g 181 t
 ORIGIN
 Query Match 26.4%; Score 306.8; DB 44; Length 552;
 Best Local Similarity 83.4%; Pred. No. 3.5e-72;
 Matches 372; Conservative 0; Mismatches 72; Indels 2; Gaps 2;
 QY 669 gaccgtgtgccaccacaaattcataagc-atactgcatcgagcaattga 727
 Db 550 GCCGGGATGCGCCACCATCAGAAATCTCATTCGCCATACGCGATCGTCNCAATTGTT 491
 QY 728 gttgaaagcagttgattc-agatacctcaactctggaactgtttatcagc 786
 Db 490 GTTGCAATGAGTGTGATTCGCGAAACTCAAGTTCAGCGCTATCTTATCAGC 431
 QY 787 ttcatgtaaagcatttactgataatgcagatcaactgtgtaacataatgagt 846
 Db 430 TTCATGTAAAGCATTTCTATTGCTGATGCTGATCACTTTGTCACAAACAATGATG 371

QY 847 aaacgaatgacaacagttgtttacaaaaataccacgtgattcagagatcgtgac 906
 Db 370 GAATAGATMGACAACTGATGTCACAAAAATACCAGCTGATTTAAGCAACGTGACG 311
 QY 907 cgcacattgcatgaatcaaatggaagtgaattccgcgaacttgaagtcacattc 966
 Db 310 TGCATCATTTACAGCAATCAACGGAAGTTCAATTCGGCTGACCTCCACCTTCATC 251
 QY 967 attagcgaaccttctgataaagaccttcacataaacatgctctgaaagcatt 1026
 Db 250 ATTGCGCATCTCGCTGATGCTATAGCCCTTTCAGTATCAATTCCTTCAAAATGATT 191
 QY 1027 acgtgatataaatgattgtattgaagttggttcacatcaattctggtcaccctc 1086
 Db 190 ACGATATTAAAGTTGACAGATTAATGATGTTGGACACATTCATTCTGTTACCTTC 131
 QY 1087 gacacattgtcaaccggaatcgaaat 1112
 Db 130 TACACGTTATCAACCGGATGACAT 105
 RESULT 9
 AI317885/c 615 bp mRNA EST 17-DEC-1998
 LOCUS SMOV3MCAM06H12SK Onchocerca volvulus molting L3 larva cDNA
 DEFINITION (SL96MLM-Ovml3) Onchocerca volvulus cDNA clone SMOV3MCAM06H12 5', mRNA sequence.
 ACCESSION AI317885
 VERSION AI317885.1 GI:4033152
 KEYWORDS EST.
 SOURCE Onchocerca volvulus.
 ORGANISM Onchocerca volvulus.
 Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida; Filarioidea; Onchocercidae; Onchocerca.
 REFERENCE 1 (bases 1 to 615)
 AUTHORS Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
 TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
 JOURNAL Unpublished (1997)
 COMMENT Contact: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genome@smith.edu
 Seq primer: pBluescript SK.
 FEATURES
 source
 1..615
 /organism="Onchocerca volvulus"
 /strain="Kumba, Cameroons"
 /db_xref="taxon:6282"
 /clone="SMOV3MCAM06H12"
 /clone_1lb="Onchocerca volvulus molting L3 larva cDNA (SL96MLM-Ovml3)"
 /dev_stage="molting L3"
 /lab_host="XLI-Blue MRF"
 /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (forest strain). The L3 were cultured in 20% FCS in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volvulus molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda Uni-Zap XR vector and has 1 x 10E6 independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustigm@mc.org)."
 BASE COUNT 154 a 106 c 111 g 181 t
 ORIGIN

AA625020/c 363 bp mRNA EST 12-NOV-1997
 LOCUS SMOV3MCA691SK Onchocerca volvulus molting L3 larva cDNA
 DEFINITION (SL96MLW-Ovml3) Onchocerca volvulus cDNA clone SML3C0691 5', mRNA
 sequence.
 ACCESION AA625020
 VERSION AA625020.1 GI:2537422
 KEYWORDS EST.
 SOURCE Onchocerca volvulus.
 ORGANISM Onchocerca volvulus.
 Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;
 Filarioidea; Onchocercidae; Onchocerca.
 REFERENCE 1 (bases 1 to 363)
 AUTHORS Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
 TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
 JOURNAL Unpublished (1997)
 COMMENT On May 9, 1995 this sequence version replaced gi:802401.
 Contact: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genome@smith.edu
 Seq primer: PBluescript SK.
 Location/Qualifiers

FEATURES

1..363
 /organism="Onchocerca volvulus"
 /strain="Kumba, Cameroons"
 /db_xref="taxon:6282"
 /clone="SML3C0691"
 /clone_lib="Onchocerca volvulus molting L3 larva cDNA
 (SL96MLW-Ovml3)"
 /dex_stage="molting L3"
 /lab_host="XLI-Blue MRF"
 /note="Vector: Lambda Uni-Zap XR; Site_1: Eco RI; Site_2:
 Xho I; Filarial nematode parasite of humans. Third-stage
 larvae, L3, were isolated from infected black flies in
 Cameroon (forest strain). The L3 were cultured in 20% FCS
 in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
 culture. L3 of O. volvulus molt to fourth-stage larvae by
 day 5 in culture. mRNA was isolated from approximately
 6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3
 in culture, and converted to double-stranded cDNA using
 reverse transcriptase and oligo(dT) followed by RNase H
 and DNA pol I. The library was constructed in the lambda
 Uni-Zap XR vector and has 1 x 10E6 independent
 recombinants and the average insert size is ~1200 bp. The
 library was constructed by Sara Lustigman and Michelle
 Lizotte-Waniewski in the laboratory of Dr. S. A. Williams.
 The library is available from Dr. Sara Lustigman (email:
 slustig@nbc.org)."
 BASE COUNT 104 a 71 c 110 t 1 others
 ORIGIN

Query Match 24 0%; Score 278.8; DB 36; Length 363;
 Best Local Similarity 85.4%; Pred. No. 1,1e-64;
 Matches 310; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

OY 410 tgatgaagaagctgacgcagatccgacatattgtatagcttcttgccagccatt 469
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 363 TCGTGAACAGGTGTGATCGATCCGATANTGTGAGACGTGCTTCTTGCCAGCCATT 304
 OY 470 aaatctgttgatattccaatatttagcaaatattatcaagagcaatcatcagca 529
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 303 AAATAGTGTGATATTCACAGTTGTTGAGCAAAATATTGTTCCAAAGCACACATCAGCA 244
 OY 530 tttagaatttccagagcttaccgttaccatcatgcacaaagcaagatgagcaaccg 589
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 243 TTTAATTAATTCACCTTATACCGTTTACCATCATCAACAGCAATGAATGTACACAGCA 184

OY 500 cagaagaatlatcaacggttccagaatcgcattgcatltagatlaaacgtgctgaccaatg 649
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 183 CAGATGTATCAACTGTTTACGATCATCATGTCATTTATAGTTAGACGTGTGACCATG 124
 OY 650 atagaagaattgacgtgttgacccgttgcaccatccaaatttcaaggagactat 709
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 123 GTAGCAAAATGTGAACAGGTTGTCGATGCATCCATCAAGATCTCATATATGGCATACA 64
 OY 710 ggcacgcgacacatttgattgaaagcaagtgtgattgagatcacctcaatcgttca 769
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 63 GGCATCGGTACAAATTGTGTTTCCAAATGCAATGCAATTTCATTTCCGAACATTCGAAGTTGAGCG 4
 OY 770 cta 772
 ||||
 DB 3 CTA 1

RESULT 12
 AA841200/c 354 bp mRNA EST 02-MAR-1998
 LOCUS MB3D6AA4G10T3 Brugia malayi day 6 post-infection third stage larvae
 DEFINITION SAG96MLW-Bml3d6 Brugia malayi cDNA clone 3D6AA4G10 5', mRNA
 sequence.

ACCESION AA841200
 VERSION AA841200.1 GI:2922536
 KEYWORDS EST.
 SOURCE Brugia malayi.
 ORGANISM Brugia malayi.
 Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;
 Filarioidea; Onchocercidae; Brugia.

REFERENCE 1 (bases 1 to 354)
 AUTHORS Blaxter,M.L., Waterfall,M., Daub,J., Lizotte-Waniewski,M., Baron,L.
 and Jones,S.J.
 TITLE Genes expressed in day six post-infection, third stage larvae of
 Brugia malayi

JOURNAL Unpublished (1997)
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2285052.
 Contact: Blaxter ML
 Institute of Cell, Animal and Population Biology
 University of Edinburgh
 Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
 3JF, UK.
 Tel: +44 131 650 6760
 Fax: +44 131 670 5450
 Email: mark.blaxter@ed.ac.uk
 The ABI trace of this sequence can be viewed at
<http://www.sanger.ac.uk/brugia/3d6/MB3D6AA4G10T3.html>
 Seq primer: T3.

FEATURES

1..354
 Location/Qualifiers
 /organism="Brugia malayi"
 /strain="TRS Labs"
 /db_xref="taxon:6279"
 /clone="3D6AA4G10"
 /clone_lib="Brugia malayi day 6 post-infection third stage
 larvae SAG96MLW-Bml3d6"
 /sex="mixed"

/dex_stage="third stage larvae, six days after infection"
 /lab_host="E. coli XLI-Blue"
 /note="Vector: lambdaZapRII (UniZap XR); Site_1: Eco R I
 (5' end); Site_2: Xho I (3' end); Brugia malayi is a
 lymphatic filarial nematode parasite of humans. mRNA was
 prepared from third stage larvae of Brugia malayi isolated
 from the peritoneal cavity of jirds six days after
 infection. The mRNA was converted to double stranded cDNA
 using reverse transcriptase and oligo(dT) followed by
 RNase H and DNApol I. The library had 2 x 10E5 independent
 recombinants and average insert size was 900 base pairs.
 The library was constructed by Michelle Lizotte-Waniewski.
 The library is available from The Filarial Genome Project
 Resource Center: contact Dr. S.A. Williams, Clark Science
 Center, Smith College, Northampton, MA 01063 USA phone +1
 413 585 3826 fax +1 413 585 3786 email genome@smith.edu."

BASE COUNT 105 a 81 c 84 g 84 t
ORIGIN

Query Match 23.2%, Score 269.2; DB 39; Length 354;
Best Local Similarity 85.0%; Pred. No. 4.2e-62;
Matches 301; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Oy 171 gtaatcaacgagcaaaacttgattatcatcgcctaactcaaggtgtgtatcatc 230
Db 354 GCAATGCACTAGTGTAGACTTGATTTCTTCATGATATTCAAGGCGCTGATATCAGTCC 295
Oy 231 gtacatcaatgatattctccggttcgcagacbtcttcttgagtaaacgaagtgcag 290
Db 294 TCACGTGACGACGCTTCTGTGCTCCGACCTTTCTTGAGCAAGCGGACCTGAGCAG 235
Oy 291 ctgcagcaggtttgtctgcgcacacactgttttaacagctccgaatcctgtgtctg 350
Db 234 CTGCTGCAAGTGTAACTCGGACCTCCCGTTTACAGCTCCAAATCTTGCGGCTCTG 175
Oy 351 aacattgtgtcgaacacatctgcctatttggttcttaagttaactgactgacatt 410
Db 174 AACATTGTGCGCGACACATTCACGTGTTGTTCTTTAATAGTACTGATCTGGCATT 115
Oy 411 gatagaaagcttgatcgcacgcataattgtatacgtgaagctcttcggcagcactta 470
Db 114 GATGAAAGATGGCATGCGATCCGATTTGTACACATGCGCTTCTTGACGACGCACTAA 55
Oy 471 aatcgtgtgatatccaaattatttaagcaaatatttcaagagacacatccat 524
Db 54 GATCGGTGATATTCAAATTTGTAGCAAAATATTGTGACGAGACCAACCAT 1

RESULT 13
AA625024/c 466 bp mRNA EST 12-NOV-1997
LOCUS SMOV3MCA158SK Onchocerca volvulus molting L3 larva cDNA
DEFINITION (SL96MLM-Ovml3) Onchocerca volvulus cDNA clone SML3CO758 5', mRNA
sequence.

ACCESSION AA625024
VERSION AA625024
KEYWORDS GI:2537426
SOURCE EST.

ORGANISM Onchocerca volvulus.
Onchocerca volvulus

Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;
Filarioidea; Onchocercidae; Onchocerca.

REFERENCE 1 (bases 1 to 466)
AUTHORS Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL Unpublished (1997)
COMMENT On May 9, 1995 this sequence version replaced gi:802407.
Contact: Steven A. Williams

Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu

FEATURES
source
Seq primer: Bluescript SK.
Location/Qualifiers
1..466

/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SML3CO758"
/clone_1id="Onchocerca volvulus molting L3 larva cDNA
(SL96MLM-Ovml3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in

Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dt) followed by RNase H
and DNA pol I. The library was constructed in the lambda
uni-zap XR vector and has 1 x 10⁶ independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Waniewski in the laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustig@embyc.org)."

BASE COUNT 126 a 91 c 93 g 144 t 12 others
ORIGIN

Query Match 23.1%, Score 267.8; DB 36; Length 466;
Best Local Similarity 80.7%; Pred. No. 1.1e-61;
Matches 305; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Oy 735 aagcagtgatcattcagatcctcaatctgtgcactcaactggtttatcagcttccatg 794
Db 466 ATGCAGTTGTCTATTCGGAAGACGTCNAGTGAGCCCTAGCTGTCTTATCACCCTTCACATGT 407
Oy 795 aaaaagctgtacatgcagatcacagatcaacttggtagcaaatatgatgaaacgaa 854
Db 406 AAGAGCANNTTGATTTGATGATATGCTCGATCAACTTTTGTACAAANCAATGAGATGAGATGAGA 347
Oy 855 tgacacagctgtgtgtltaaaaaataccacgltgataltaagagatcgtltaacgcaaat 914
Db 346 TGACAAACAAGTGTGATGACCAACAATACACGCGANNNTAAGCAAGGTGTAGGTGTACAT 287
Oy 915 tgcattgaatcaaatggaagtgaatctccgcaacttgacgtccactcatcatcagtcg 974
Db 286 TACAGGAATCAANCGAAGGTCAATTCGCGNACCTGACGTCACACCTTCATCATTTTCCGC 227
Oy 975 aacctctgtatcaaaagaccttcacataaacaatgaccttgatgcatcattgatt 1034
Db 226 ATCTGCGCTGATCGATTAAGCTTTTCACGTANANATGTCCTTCAATGATATTACAGATAT 167
Oy 1035 taaatgtatgttatttgaagtgtgtccacatcaatctctggtcaccctcgacacat 1094
Db 166 TAAAGTTGACAGTAAATGATGTTGGACCACATTCATTTGTTGTTGACCTTCTACACCGT 107
Oy 1095 tgtcaaccggaatggaat 1112
Db 106 TATCAACCGGAATGACAT 89

RESULT 14
AA618952/c 459 bp mRNA EST 12-NOV-1997
LOCUS SMOV3MCA2020SK Onchocerca volvulus molting L3 larva cDNA
DEFINITION (SL96MLM-Ovml3) Onchocerca volvulus cDNA clone SML3CO2020 5', mRNA
sequence.

ACCESSION AA618952
VERSION AA618952
KEYWORDS GI:2522828
SOURCE EST.

ORGANISM Onchocerca volvulus.
Onchocerca volvulus

Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;
Filarioidea; Onchocercidae; Onchocerca.

REFERENCE 1 (bases 1 to 459)
AUTHORS Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1405277.
Contact: Steven A. Williams

Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith

College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786

FEATURES

Location/Qualifie

```

/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="Sml3CO2020"
/clone.lib="Onchocerca volvulus molting L3 larva cDNA
(SL96MLM-OvML3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MR"
/notes="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ml3)/2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by Rnase H
and DNA pol I. The library was constructed in the lambda
Uni-ZAP XR vector and has 1 x 10E6 independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams
The library is available from Dr. Sara Lustigman (email:
slustig@nbc.org)"

```

BASE COUNT	131 a	85 c	91 g	150 t	2 others
ORIGIN					

Query Match	22.7%	Score 263	DB 36	Length 459
Best Local Similarity	83.9%	Pred. No. 2.1e-60		
Matches 296	Conservative	0	Mismatches 57	Indels 0
				Gaps 0

Qy	760	aatctgagcaactaacggttttaccagctccatcgtataaagaacttaccgatacgcagc	819
Db	433	AAGTGGAGCCCTACGTCTCTATCACCTTCCAGTMAAAGCATTTGATTCGATATGCTCG	374
Qy	820	atcaactctggtataacaataatacgtatgataaagcaaatcgcacaacgctgtgtgtatacaaat	879
Db	373	ATCAACTTTTGTGCACAAACATGAGATGGATGGATGAGACAAACAGTGTGCACCAAAAT	314
Qy	880	acacacgtgatactagaagatcgtgtacgcgcaaatctgccttgatacaaatcgtgaagtgaat	939
Db	313	ACCACGTGGATTATACCAACAGTGTACGTGTACATTACACGGAATCAACGGAACTTCAAT	254
Qy	940	tccgcgaactctgacgtccacactcattcaatcagcgaaccttcctgtatacaataagaacctt	999
Db	253	TCCGGCTACCTGCAGTCACACCTTCATCATTTTCGGCATCTCGCTGATCTATMAACCTTT	194
Qy	1000	caactaaacaatgcacttcgagatcgtatcaatcagtgattaaataatcgtatcgtatcgtgaagtctg	1055
Db	193	CAGCTATACANNGTCCTTCAATATGAGATTACAGAGATTTAAAGTTGACAGTATATGATGCTTGG	134
Qy	1060	tccacatcattatctcgtctcactccttcgacaccattgtcaccocggaatcgaat	1112
Db	133	ACCACATTCACATTTTCGTTCCACCTTCACACCGTATACACCGGAATGGCAT	81

RESULT	15
AA625010/c	
LOCUS	395 bp mRNA
DEFINITION	SMOVMCA318K Onchocerca volvulus molting L3 larva cDNA (SI95MELW-OvM13) Onchocerca volvulus cDNA clone SMML3CO318 5', sequence.
ACCESSION	AA625010
VERSION	AA625010.1 GI:2537412

KEYWORDS	EST.
SOURCE	Onchocerca volvulus.
ORGANISM	Onchocerca volvulus

REFERENCE 1 (bases 1 to 395)
AUTHORS Williams, S.A., Iizotte-Waniewski, M., Laney, S. and Lustigman, S.
TITLE Genes expressed in molting L3 larvae of *Onchocerca volvulus*
JOURNAL Unpublished (1997)
COMMENT On May 9, 1995 this sequence version replaced gi:802391.

Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel.: 413/5853826
 Fax: 413/5853786
 Email: genome@smith.edu
 Seq primer: pb1nucscript SK.
 Location/Qualifiers
 1..395

JRES	Location/Qualifiers
source	1. .395

```

/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMML3C0318"
/clone_1fb="Onchocerca volvulus molting L3 larva cDNA
(S169MDM-OvmL3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site.1: Eco RI; Site.2:
Xho I; Vectorial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (mL3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNA pol I. The library was constructed in the Lambda
Uni-Zap XR vector and has 1 x 10E6 independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Maniewski in the laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustigm@nbc.org)."

```

BASE COUNT	118 a	76 c	88 g	112 t	1 others
ORIGIN					
Query Match		22.4%	Score 260.4;	DB 36;	Length 395;
Best Local Similarity		84.0%;	Pred. No. 1e-59;	57;	Indels 1;
Matches 305; Conservative		0;	Mismatches 57;	Indels 1;	Gaps 1.

QY	296	gcgaagtttgcgtgcgagccacccgctttcaacgcttcgaactctcttgatctcgcaacat	355
Db	375	GCAGCTGCAGACACCAACCATTTTCGTGANGAACCTCTCCAAATCTTGTGGTTCGCACAT	316
QY	356	tgtgtgcgaacacatctcgctatcttgatctctttaaigtgtaactgactgcgcatctga	415
Db	315	TGTGTGCTGCACATTCACCTGTTGGTTCCTTAATTGTAATCAATTTGCATTTGGTAG	256
QY	416	aaagcctgtatcgatccgcgaatttttaacgtagaactctcttgcaagcattaaatct	475
Db	255	AACGTTGTGATTCATCCGATATTGTGAGACGAGCTCTTGTCCAGCATTAATAATCA	196
QY	476	gttgatatacttcaaatattatlaagaanaatattatcaagagacatccatcaacttga	535
Db	195	GTTGGATATTCCAAAGTTGTTGACGAATATTTGTCCAAAGACAAACATCGACTTTAAT	136
QY	536	attccacagatataccggttaaccatctcagacaagaacaggaatgagacaacccgcgagaa	595
Db	135	AATTCCACCTTATACCGTTTACCATCATCAACGAAGAAGATGTAATGTATACAAAGCACAAAT	76

Qy	596	gtatcaacggttcagaatcgcatgcatcttaatga taaactgctgacccaatgatagca	655
Db	75	GTATCAACTGTTTCAGATCAGATGTCATTTATGCTACACTCG-TGACCAATGCTAGCA	17
Qy	656	aat	658
Db	16	AAT	14

Search completed: April 16, 2000, 03:09:21
 Job time: 6205 sec

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 4, 2003, 09:12:27 ; Search time 1 seconds
(without alignments)
4.131 Million cell up

Title: us-10-054-562a-3
 Perfect score: 1161
 Sequence: 1 atgatgattcgtcttatgc.....ttcgtccaaatcgaaggca 1161

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 0.0

Searched: 1 seqs, 1779 residues

Total number of hits satisfying chosen parameters: 2

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 1 su

Database : us-10-054-562a-1:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Match	Length	ID	Description
	1	1161	100.0	1779	1	us-10-054-562a-1

ALIGNMENTS

RESULT 1
us-10-054-562a-1

```
Query Match      100.0%; Score 1161; DB 1; Length 1779;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	1	ATGATGATTCGCTCTATTGCTTCTGTA	CTACTACACTTATTCGATTCCTTATTCGATTCGG	60
Db	167	ATGATGATTCGCTCTATTGCTTCTGTA	CTACTACACTTATTCGATTCCTTATTCGATTCGG	226
QY	61	GTTGCAATGGTGTTCGAAGTGCAGCAG	AAATTTGAATGTGGACCAACTTCATATCAATC	120
Db	227	GTTGCAATGGTGTTCGAAGTGCAGCAG	AAATTTGAATGTGGACCAACTTCATATCAATC	286
QY	121	AATTTTAATACACGTAATGCATTCGA	GACACATGTTTATGTGAAGGCTCTTTATGATCAA	180
Db	287	AATTTTAATACACGTAATGCATTCGA	GACACATGTTTATGTGAAGGCTCTTTATGATCAA	346
QY	181	GAAGGTGCGCTAATGATCAAGGTGC	AGCTCAAGTTGCGGAAATTCACCTCCATTTGAT	240
Db	347	GAAGGTGCGCTAATGATCAAGGTGC	AGCTCAAGTTGCGGAAATTCACCTCCATTTGAT	406
QY	241	TCATGCAATGTTCGCGGTACAGATCT	CTGAACTCACGTGGTATTTTTGTAAACAACACT	300
Db	407	TCATGCAATGTTCGCGGTACAGATCT	CTGAACTCACGTGGTATTTTTGTAAACAACACT	466

THIS PAGE BLANK (USPTO)